NMR study of the *Escherichia coli* 70S ribosome particle using residual dipolar couplings

Xiaolin Wang

Department of Structural and Molecular Biology

2012

A dissertation submitted for the degree of Doctor of Philosophy

Abstract

The ribosome is the 2.5-MDa complex that is responsible for production of biologically active polypeptide chains. Crystallographic and cryo-EM studies have revealed the intricate motions of the ribosome during protein synthesis, but less is known about the highly dynamic regions such as L7/L12 stalk protein, which has an important role during this process. Moreover, there is currently very little structural information on the dynamic nascent polypeptide chain (NC), which during its progressive emergence from the ribosome has its first opportunity to acquire structure, folding in a co-translational manner. While the formation of tertiary structure occurs outside of the ribosomal exit tunnel while the NC remains tethered there is little understanding of how this process takes place.

NMR spectroscopy, which has the ability to report on both residue-specific structure and dynamics at atomic resolution, has previously been applied to report on mobile regions of the ribosome, such as the L7/L12 stalk protein, and more recently, to the study of co-translational folding of ribosome-nascent chain complexes (RNC). For example, chemical shift studies of RNCs of an immunoglobulin domain (ddFLN-Dom5), have revealed a dynamic RNC capable of adopting the native fold.

The motional properties of both L7/L12 and the NC therefore permit the application of NMR strategies, such as residual dipolar couplings (RDC), to derive direct three-dimensional structural information. RDCs encode detailed structural information in the form of distance-independent bond-vector orientations, and can be exploited for both structured and unstructured proteins. This thesis describes steps towards the development of RDC techniques for probing the structural and dynamic properties of RNCs. Initial application to the L7/L12 stalk protein was used as the foundation for subsequent studies of RNCs.

Alignment in phage enabled a set of RDCs to be measured for both ribosome-bound and isolated L7/L12, yielding the first direct structural information on the L7/L12 protein in its ribosome-bound form.

The application of RDCs to the study of RNCs required extensive development of the strategy for the production of isotopically-labelled RNCs, which resulted in significant improvements in both the yield and quality of the final RNC samples. The RDC study of a model RNC, Dom5+110, highlighted the inherent challenges associated with the finite sample lifetime, low sample concentration and restricted mobility of the RNC. These were addressed with the study of the intrinsically disordered αSyn-RNC, which enabled the measurement of N-H RDCs. A potential interaction between the αSyn NC and the ribosome prevented acquisition of RDCs for all residues, but further studies are underway to improve the quality of the spectra with a view to obtain the complete set of RDC data. Together these data provide a solid foundation for an RDC methodology for application to the ribosome and RNCs, and will enable more detailed structural and dynamical studies of the emerging nascent chain.

Preface

This thesis describes research undertaken at the Department of Structural and Molecular Biology, University College London between October 2008 and September 2012. The dissertation is the result of my own work and includes nothing that is the outcome of work done in collaboration except where specifically indicated in the text.

Acknowledgements

Firstly, I would like to express my gratitude to my supervisor, Dr John Christodoulou, for giving me the opportunity to undertake my PhD studies and for his supervision, support and advice throughout my PhD. I am grateful to my thesis examiners Dr Jason Schnell and Dr Snezana Djordjevic who have taken the time to edit my thesis with great details, and for their helpful suggestions.

Many thanks to Dr John Kirkpatrick, who has always been supportive and helpful in every respect. I would like to thank him for improving my understanding of my biomolecular NMR knowledge, guiding in performing NMR experiments and for his editorial advice and proof-reading of my thesis.

I would also like to thank all members of the JC group. Dr Lisa Cabrita has been very helpful in the biochemistry work, which is a very important part of this thesis. I would like to thank Dr Helene Launay and Maria-Evangelia Karyadi with whom I worked closely and had lots of support from, as well as Annika Weise, Anne Wentink, Geraldine Levy, Luke Goodsell, Toshitaka Tajima and Dr Chris Waudby. Annika has been very friendly and supportive during the whole time we worked together.

I would like to thank my cousin for looking after me during my PhD and John for all his support and patience all the time, I would not have made it though without him.

Finally, and most importantly, I want to thank my parents, who have been a constant source of support and encouragement, without which I would certainly not have come this far.

Contents

List of Figures

List of Tables

Abbreviations

ATP	Adenosine triphosphate
AUC	Analytical ultracentrifugation
BME	β-mercapto ethanol
CPD	Composite-pulse decoupling
CPMG	Carr-Purcell-Meiboom-Gill
CRINEPT	Cross-correlated relaxation-enhanced polarisation transfer
Cryo-EM	Cryo-electron microscopy
CTD	Carboxyl terminal domain
DNA	Deoxyribonucleic acid
EF-G	Elongation factor G
EF-Tu	Elongation factor Tu
FRET	Förster resonance energy transfer
GAR	GTPase-associated region
GTP	Guanosine triphosphate
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulphonic acid
HSQC	Heteronuclear single quantum correlation
IDP	Intrinsically disordered protein
IF	Initiation factor
INEPT	Insensitive nuclei enhanced by polarisation transfer
IPAP	in-phase/anti-phase
IPTG	Isopropyl-β-D-thiogalactopyranoside
iRED	reorientational eigenmode dynamics
kDa	kilo Dalton
LBT	Lanthanide binding tag
MC	Monte Carlo
MD	Molecular dynamics
MDa	mega Dalton
MES	2-(N-morpholino)ethanesulphonic acid
mRNA	messenger RNA
NOE	Nuclear Overhauser effect
NTD	Amino terminal domain
OD_{600}	Optical density at 600 nm
PAGE	Polyacrylamide gel electrophoresis
PCS	Pseudocontact shift
PD	Parkinson's disease
psi	Pound-force per square inch
PTC	Peptidyl transferase center

RDC	Residual dipolar coupling
RNA	Ribonucleic acid
RNC	Ribosome-nascent chain
RF	Release factor
RMS	Root mean square
RMSD	Root mean square deviation
RPM	Revolutions per minute
SDS	Sodium dodecyl sulphate
Semi-CT	Semi-constant-time
Tico	Tight-coupled
tRNA	transfer RNA
TROSY	Transverse relaxation-optimised spectroscopy
Tris	Tris(hydroxymethyl)aminomethane

Chapter 1

Introduction

1.1 Introduction to the structure of the *Escherichia coli* 70S ribosome

The ribosome is present in all kingdoms of life and is the molecular machine that enables protein synthesis to take place. The 2.5-MDa ribosomal complex in *Escherichia coli* is a 70S particle which is composed of 56 proteins[1] and 3 large RNA molecules distributed between two subunits, the 50S and 30S. The proteins L1-L36, together with the rRNA molecules 5S and 23S, form the 50S subunit, while the proteins S1-S22, together with the 16S rRNA, comprise the 30S subunit. The RNA content represents about two-thirds of the total mass of the ribosome (see Table 1.1) and is also the main constituent of the functional core of the ribosome, the peptidyl transferase centre (PTC) (Figure 1.1 B), which is shared between the 30S and 50S subunits.

It is well known that the function of the ribosome is to translate genetic information contained in the incoming mRNA transcript into functional proteins [2], but exactly how the ribosome performs its function was not well understood until very recently, when high-resolution structures of the ribosome particle, solved using methods such as X-ray crystallography [2–9] and cryo-EM [10–16] began to emerge. It is now known that the

[1] L7 is the acetylated form of L12, L8 is the complex of L7/L12 and L10, L26 and S20 are same protein and L31 has two forms [1].

Table 1.1: Components of *E. coli* ribosomes

	Ribosome	Small subunit	Large subunit
Sedmentation coefficient	70S	30S	50S
Mass (kDa)	2524	930	1594
RNA		16S	23S, 5S
RNA mass (kDa)	1664	560	1104
RNA propotion	66%	60%	69%
Protein		22 proteins	35 proteins
Protein mass (kDa)	860	370	490
Protein propotion	34%	40%	31%

ribosome functions as a ribozyme, with the catalysis of peptide bond formation driven by the PTC, which is composed entirely of RNA [17].

Moreover, these structures have revealed the details of the intricate mechanistic characteristics of ribosome function and have uncovered the method by which the 30S and 50S subunits work in unison during the translation process. For instance, the small ribosomal subunit (30S) is thought to be responsible for mRNA recognition and mediates decoding of the mRNA, whose sequence determines the polypeptide chain that is produced, by ensuring that correct base pairing takes place between the mRNA and the cognate tRNAs [18]. In contrast, the large ribosomal subunit (50S), which harbours a significant proportion of the PTC [19], is largely responsible for the peptide-bond formation that enables the progressive assembly of the nascent polypeptide chain to take place. Within the 50S subunit is another important feature associated with protein synthesis, the ribosomal exit tunnel. This is a 100 Å-long channel that extends through the length of the 50S subunit and varies in width from 10 to 20 Å. It is lined with parts of the 23S rRNA and with a number of ribosomal proteins (Figure 1.1 B and Figure 1.6). Approximately 80 Å from the PTC, the ribosome exit tunnel is at its widest, and this region is referred to as the 'vestibule' or 'exit port' [20]. Following synthesis, it is via this channel that the nascent chain (NC) makes its way through the ribosome and into the cellular milieu.

Another structural feature that is pertinent to both the 50S subunit and to protein translation alike is a set of structural elements which are collectively referred to as the GTPase-associated region (GAR) and which are highly conserved in both prokaryotes and

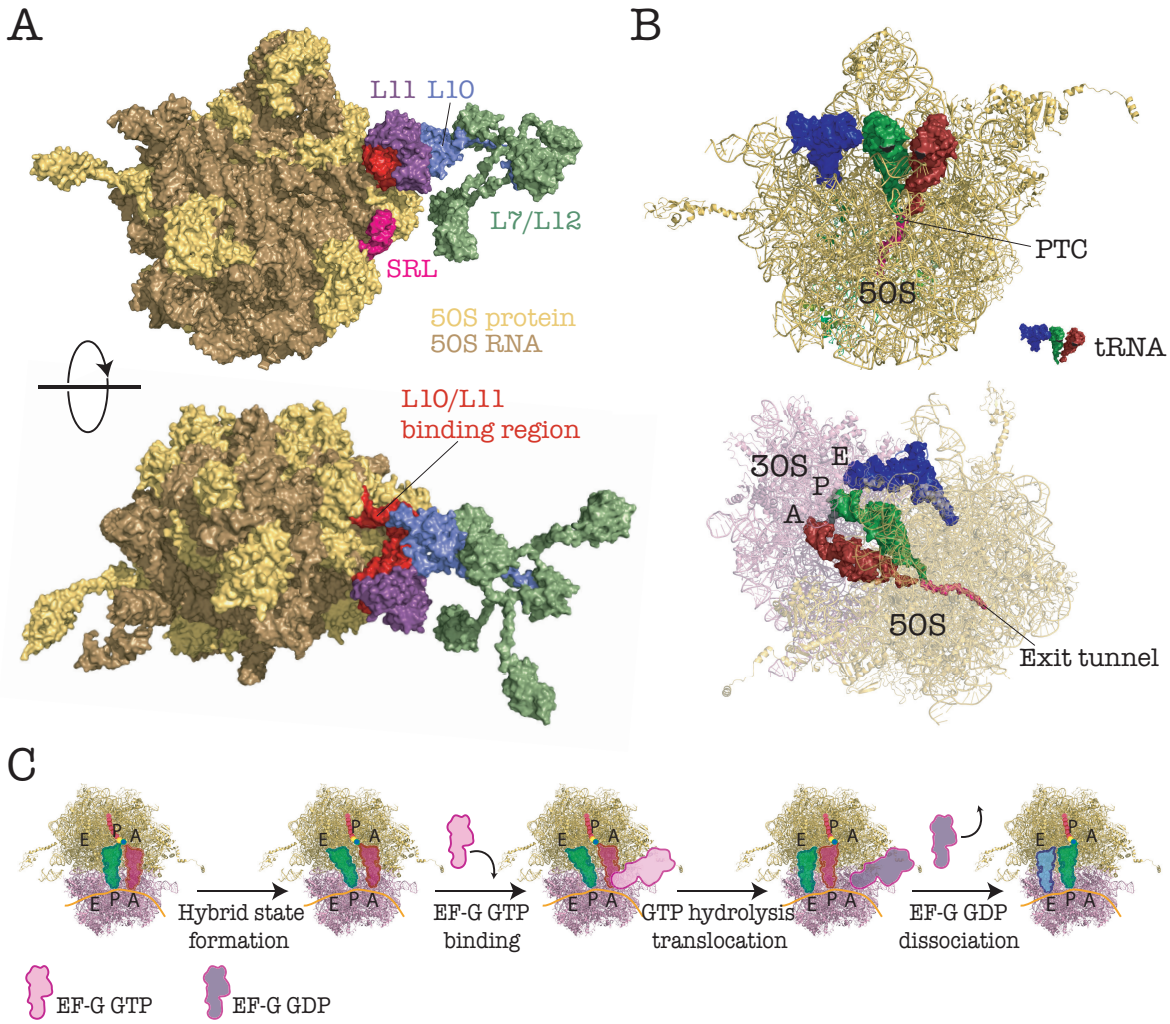

Figure 1.1: Representation of the 70S ribosome complex with the 50S (yellow, PDB 2AW4) and the 30S (pink, PDB 2AVY) subunits [21]. A: Front and top view of the 50S subunit, the components of the GTPase-associated region (GAR) are highlighted, including the sarcin-ricin loop (SRL) (23S RNA, helix 95) in magenta, the L10/L11 binding region (23S RNA, helices 42-44) in red, protein L11 in purple, L10 in blue and L7/L12 (PDB 1RQU [22]) in green. The L7/L12 and L10 proteins were modelled onto the structure by alignment via the crystal structure of the *Thermus thermophilus* 50S ribosome, PDB 3I8I [23]. B: The A-, P- and E-site tRNAs are shown in red, green and dark blue, respectively (aligned with the crystal structure of the *Thermus thermophilus* 70S ribosome, PDB 1GIX [24]). To show the position of the peptidyl transferase center, only the 50S subunit and the tRNA are shown in the top panel. The location of the ribosomal exit tunnel is highlighted by a short nascent polypeptide in magenta (aligned with the cryo-EM structure of the TnaC stalling sequence in the *E. coli* ribosome, PDB 2WWQ, [25]. C: Schematic showing the translocation of tRNA from A- and P-site to P- and E-site after peptide bond formation; GTP hydrolysis by EF-G provides energy for the movement [26].

eukaryotes [27–29]. It includes helices of the 23S rRNA, specifically helices 42-44 (L10/L12 binding loop) and helix 95 (sarcin-ricin loop), together with the large ribosomal proteins L10, L11 and L7/L12 (Figure 1.1 A). This region is so named to reflect its role in both the recruitment and the stimulation of the GTPase activity of several auxiliary factors associated with the key steps of protein synthesis: initiation (initiation factor 2, IF2), elongation (elongation factors Tu (EF-Tu), and G (EF-G)) and termination (release factor 3, RF3).

1.1.1 The process of protein translation in prokaryotes (*E.coli*)

The pioneering studies of the ribosome have also captured in exquisite detail the different stages of the translation process, which are yielding unprecedented insights into protein synthesis at an atomic level [26, 30, 31]. The process of translation to generate a polypeptide chain is composed of three stages: initiation, elongation and termination.

The initiation step begins with the initiation factors recruiting the small ribosomal subunit and facilitating its recognition of the mRNA at the translational initiation site. In prokaryotes, the first amino-acyl tRNA, tRNA-fMet, is delivered to the P-site and positioned on the start codon, which is followed closely by the recruitment of the 50S subunit by IF2, completing the formation of the initiation complex. Following delivery of the second tRNA to the A-site, the process of elongation is begun.

During elongation, amino-acylated tRNAs that complement the mRNA coding sequence are delivered to A-site of the ribosome with the assistance of EF-Tu. The recognition of the cognate tRNA by the mRNA triggers the hydrolysis of GTP within EF-Tu, which provides the energy for the amino-acylated tRNA to correctly accommodate within the A-site. This is followed quickly by the formation of a peptide bond to the amino acid present at the end of the growing nascent chain within the P-site (Figure 1.1 B), which extends into the ribosomal exit tunnel. Following peptide bond formation, a translocation step occurs, which results in progression of the ribosome complex along the mRNA from the 5' end to the 3' end. During this process, the A-site tRNA, which is attached to the growing nascent chain transfers to the P-site, and the deacylated tRNA present within the P-site is translocated to

the E-site, and subsequently expelled from the ribosome. This enables the A-site to become vacant, allowing accommodation of an incoming amino-acylated tRNA. The translocation of the tRNAs in the P/P and A/A sites to the P/E and A/P sites occurs with the assistance of EF-G and GTP hydrolysis (Figure 1.1 C). When the ribosome encounters the mRNA sequence corresponding to a stop codon, the termination process is triggered; release factors are recruited at the A-site and catalyse the hydrolysis of the ester bond in peptidyl-rRNA and release of the nascent polypeptide chain from the ribosome together with dissociation of the subunits and mRNA.

1.1.2 The dynamic properties of the ribosomal complex

Although the many crystal structures that have emerged present static representations of the ribosome, the complex itself and the process of translation are highly dynamic [32]. Cryo-EM structures, which can offer some dynamic details, together with the crystal structures, have begun to shed light on the role of dynamics, particularly in the stages involving tRNA recognition and translocation [33].

During the process of elongation, the amino-acyl tRNA (aa-tRNA) is selected by the ribosome through an interaction between the L7/L12 stalk region on the 50S ribosome and EF-Tu (Figure 1.1 A & Figure 1.2 A). The L7/L12 stalk is a highly dynamic region which contains four copies of the L7/L12 protein (six copies in eukaryotes and archaea). The C-terminal domain (CTD) of L7/L12 is very flexible relative to the core of the ribosome, being connected to the ribosome-bound N-terminal domain by a 30-residue linker. The flexibility of the L7/L12 CTD allows it to 'fish' for the ternary complex of aa-tRNA-EF-Tu-GTP, which is passively diffusing within the cellular space, and transports it to the 30S/50S interface, thereby enabling delivery of the aa-tRNA to the PTC. Despite the critical role played by the L7/L12 stalk in protein translation, its structural properties on the ribosome have not been elucidated by either X-ray crystallography or cryo-EM, which is the result of its inherently dynamic nature. NMR spectroscopy, however, has offered significant insight into both the structural and dynamic properties of ribosome-bound L7/L12 via the collection of *ca.* 100

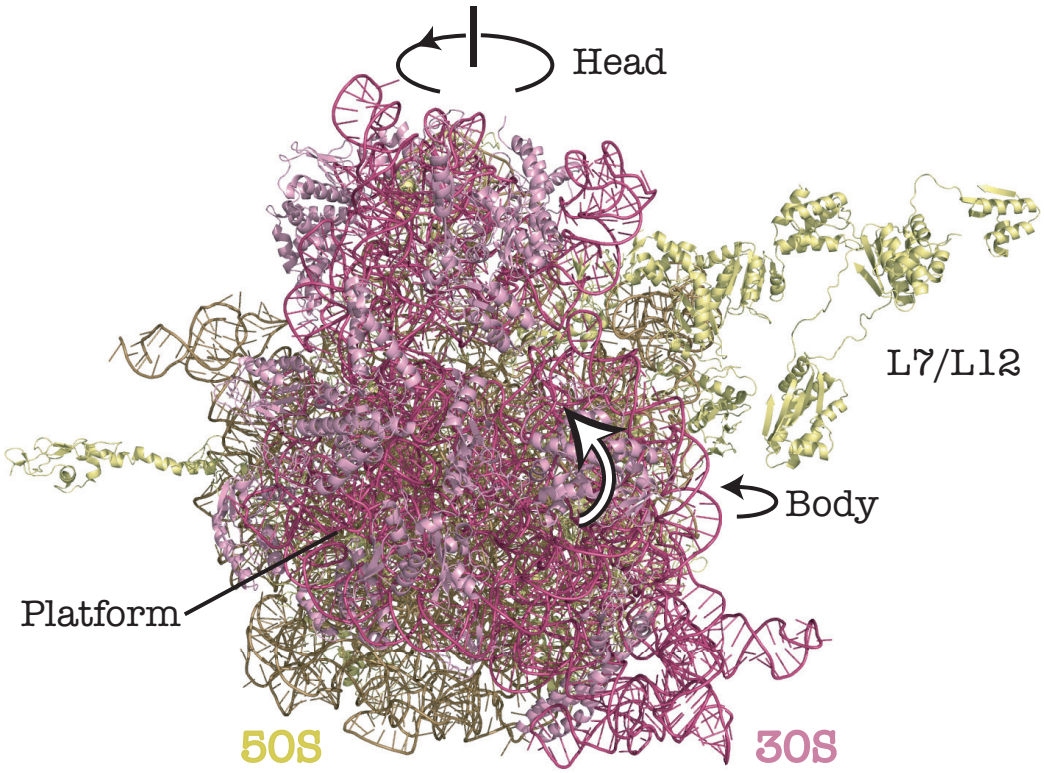

Figure 1.2: Motion of ribosome during translocation. Arrows indicate regions that move during translocation, including the head and body domain of the 30S subunit. Figure produced with PDB file 2AW4, 2AVY, 1RQU according to [2].

well-dispersed resonances that arise from residues in the CTD [34]. The very fact that it is observable by NMR indicates that the motion of the L7/L12 CTD is largely independent of the large ribosome body to which it is attached. The remainder of the 7500-residue ribosome is NMR-silent. These ideas will be explored in detail in Chapter 2.

The recognition of an incoming aa-tRNA by the ribosome has also been found to be a highly dynamic process. It involves a set of conformational changes at the decoding centre in the 30S subunit and the region responsible for GTP hydrolysis in EF-Tu [33, 35]. Upon selection of the cognate aa-tRNA to be delivered to the ribosome, the anti-codon of the tRNA is recognised by the codon of the mRNA positioned at the A-site. This recognition process induces conformational changes within EF-Tu, which results in GTP hydrolysis and, via allosteric communication, triggers the repositioning of the region surrounding the A-site so that to accommodate the cognate tRNA. Differential kinetics are observed for the transient binding of non-cognate or cognate tRNA at the A-site as observed using Förster resonance

energy transfer (FRET) studies [36]. This suggests that the ribosome can manually 'inspect' an aa-tRNA and highlights further that transient interactions are critical for correct tRNA selection.

Following accommodation of the cognate aa-tRNA at the A-site and formation of the peptide bond, the ribosome must reposition itself on the mRNA in order to increment the polypeptide chain. This involves the transfer of the acylated tRNA from the A-site to the P-site and, simultaneously, the transfer of the deacylated tRNA from the P-site to the E-site. This translocation process involves a large set of concerted conformational changes associated with the movement of the ribosome relative the mRNA transcript, which correspond to a reversible movement in which there is a counterclockwise 'ratchet-like' rotation of the 30S subunit by 8-10 ° (Figure 1.2, white arrow), and are essential for the transformation of the A/A and P/P tRNA configuration to the A/P and P/E configuration (Figure 1.1 C) [37–40]. This intact subunit movement is accompanied by an orthogonal rotation of the so-called 'head' domain of the 30S subunit to assist control of the tRNA positioning during translocation (Figure 1.2) [21, 38, 41]. In summary, the concerted movements of the P-site tRNA into the hybrid P/E configuration are the result of a set of combined rotations of the small subunit relative to the large subunit and rotation of the 30S head domain. These conformational changes are thought to be mediated to some extent by movements of the ribosomal proteins L1, L11 and L7/L12 [37–40, 42, 43] and are catalysed by the presence of EF-G.

A collection of both X-ray and cryo-EM structures [12, 14, 37, 38, 40, 43–48] correspond to a series of snapshots that represent the various stages of the ratchet-like translocational process. An inspection of the array of structures allows the tentative identification of four principal states, R_0, R_1, R_2 and R_F [2], in which the intermediate state R_2 is thought to be responsible for the formation of the tRNA hybrid state [43]. Indeed, drawing together FRET [49], cryo-EM [40] and X-ray crystallography studies illustrates that the ribosome is subject to oscillations associated with the reversible ratcheting and unratcheting movements during translocation. Together with the observation that each functional state is a hybrid

conformation, this suggests that elongation is a thermodynamically driven process [30]. Moreover, the oscillation of the ribosome between different states suggests the presence of low energetic barriers, but these are more readily overcome through the effects of GTP hydrolysis within the EFs, which drives the process of translation forwards [30, 33]. These combined studies illustrate the highly dynamic nature of translation and emphasise many of the molecular details of ribosome motion remain to be unveiled.

In addition to the physical movements of the subunits relative to each other, another region of the ribosome which is associated with dynamic behaviour and is of special interest with respect to the emerging nascent chain, is formed by the PTC and the ribosomal exit tunnel. Recent studies of ribosome-nascent chain complexes (RNCs) have provided details of the interaction between the P-site tRNA, the ribosomal exit tunnel and the nascent chain (NC) [25, 50–52], and the manner in which the ribosome can 'sense' the nascent chain. In the study of SecM-stalled RNCs, specific interactions made between residues within the SecM sequence and the ribosomal exit tunnel induce conformational changes at the A-site of the PTC which result in translational arrest [51]. Further evidence that allostery is present between the ribosomal exit tunnel and the PTC comes from cryo-EM studies of TnaC-stalled RNCs, which reveal that the set of conformational changes observed in the ribosomal RNA within the A-site were distinct from those associated with SecM-mediated translational arrest [51].

Emerging computational studies have also begun to enhance our understanding of the translational machinery. These have led to models representing the dynamic ribosome movements during translocation [53, 54] as well as probing the structural aspects of the ribosome, in particular the ribosomal exit tunnel. These studies have suggested that the ribosomal exit tunnel is a largely rigid structure with an inert inner surface, whose hydrophilicity enables NCs to move easily through the tunnel during translation [55, 56]. Moreover, these studies are increasingly being extended to the study of the NC [57–59], with models that aim to predict the conformational space that is accessible to a NC upon its emergence from the ribosome.

1.2 Introduction to protein folding and co-translational folding

1.2.1 *In vitro* protein folding: Anfinsen's theory

Proteins are essential and are the basis of all living systems from viruses to bacteria and from plants through to mammals. They have evolved a myriad of different functions, which are critical for the correct functional activity of cells including division, replication, signalling and transport. The functional aspects of a protein are closely related to its unique three-dimensional structure, whose characteristics and folding are dependent on the amino acid sequence of the polypeptide chain [60–62]. It is well-known that a protein performs its biological activity in an efficient manner only if it adopts its correct native structure; these may reflect highly-ordered assemblies, as in the case of multi-domain complexes (e.g. the molecular chaperone, GroEL [63]), a metastable native fold (e.g. the serine proteinase inhibitor, antitrypsin [64]), as well as natively unfolded ensembles (e.g. the synaptic protein, alpha-synuclein [65]).

Within the cell, most proteins exist only transiently and their cellular concentrations are finely regulated to maintain cellular homeostasis [66]. Under conditions of cellular stress or due to the influence of destabilising genetic mutations, the balance of proteins can be altered, which can result in protein misfolding. Misfolded proteins are usually degraded via quality control processes within the cell, but in some cases they can aggregate and lead to a number of different diseases, particularly those associated with neurodegeneration [67]. An understanding of protein folding is therefore not only important for understanding fundamental cellular processes, but also the basis for misfolding, and the ways in which this process can be prevented or ameliorated.

In 1973, Anfinsen demonstrated that the protein ribonuclease A could refold into an active enzyme following denaturation [61], which revealed that the native fold of a protein is governed by its amino acid sequence. In addition, it supported the thermodynamic hypothesis, which suggests that the polypeptide chain is driven to acquire a thermodynamically-

preferred conformational state which is of the lowest free energy. In principle, during the process of folding, the polypeptide chain can sample a very large number of possible conformations, (for example, a 150-amino acid protein can adopt 10^{300} conformations), however, the time required for the polypeptide to sample all the possible conformations to attain its native fold would be longer than the age of the universe, which is clearly not true, this is commonly referred to as Levinthal's paradox [68]. It is well established, that the process of folding does not proceed stochastically, and occurs on a μs-to-ms timescale. Over the last 50 years, a large number of studies, primarily *in vitro* and *in silico* using isolated proteins, have examined the mechanistic principles underlying the protein folding process. These have involved the characterisation of protein folding both under equilibrium conditions as well as the determination of folding rates [69] using a range of biophysical and structural techniques. These studies have sought to characterise the conformational states sampled by proteins during the folding process, which aside from the native and unfolded states, include the formation of transient transition states [70] as well as folding/unfolding intermediate ensembles [71, 72], which serve to guide the folding process.

Experimental strategies that are used to investigate protein folding typically involve an isolated protein and its denaturation (and renaturation) using both chemical denaturants (e.g urea, guanidium) and elevated temperatures. Biophysical approaches (*e.g.* fluorescence, circular dichroism) are employed to evaluate the thermodynamic and kinetic characteristics of the polypeptide chain as it visits many conformations during folding/unfolding. Increasingly, mechanical unfolding/folding using atomic force microscopy and molecular tweezers is being used to study folding, and present the closest mimic of a nascent chain bound to the ribosome [73]. NMR spectroscopy (discussed in Section 1.3.2), has also been used to characterise the range of conformations adopted by the polypeptide chain, and is particularly effective as it is non-destructive and can be performed under native conditions.

As a result of both experimental and theoretical approaches, a contemporary view of the process of protein folding has been developed, in which it is described in terms of a free energy landscape or 'folding funnel' (Figure 1.3). At the top of the funnel, the

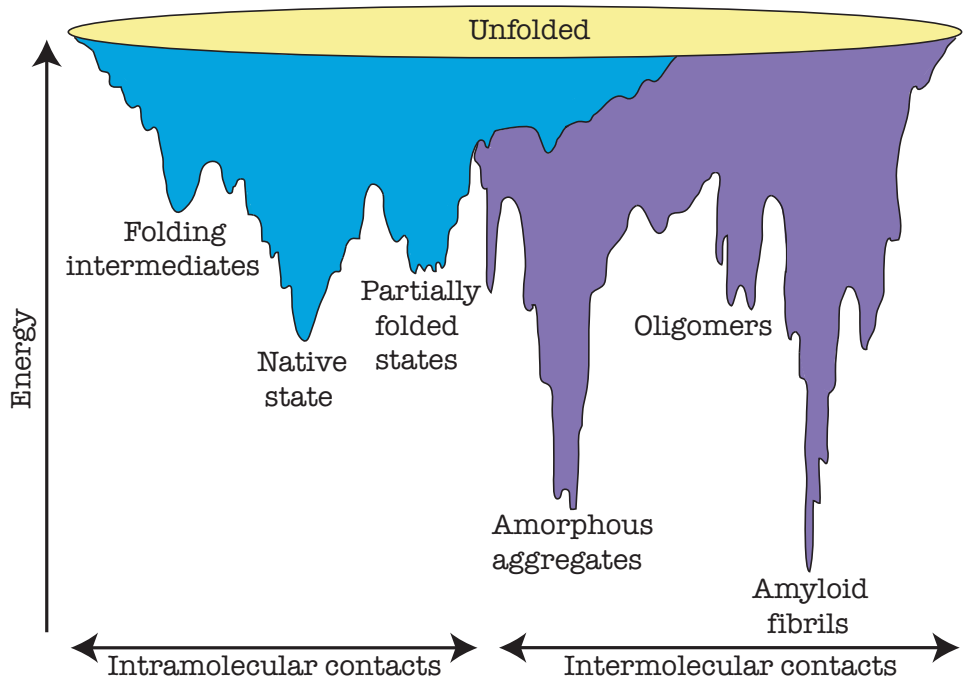

Figure 1.3: The free-energy landscape of protein folding and aggregation. The blue surface shows the multitude of conformations in moving towards native state, whereas the purple surface shows the conformations in moving towards amorphous aggregates or amyloid fibrils. Reproduced according to [74].

unfolded polypeptide chain is at its highest free energy, from which it can in principle, sample all possible conformations. During the process of folding (blue area), the formation of intramolecular contacts between amino acids begins a biased search towards the lowest free energy state (represented by a global minimum), which in nearly all cases corresponds to the biologically-active state of the protein. During the process of folding to its biologically-active state, the polypeptide chain may access obligatory or non-obligatory ensemble intermediates (represented by local minima) as well as transition-state ensembles [69]. Together these studies have suggested that a polypeptide chain will only sample a small number of possible conformations during the folding process.

Many folding studies have shown that small, single-domain proteins can fold unassisted to biologically active states under appropriate conditions [61] and within microseconds [75], whereas topologically complex or multi-domain proteins can take up to many seconds to fold [76] and in some instances require the assistance of molecular chaperones. It has also been found that the rate of folding is related to the number of local contacts present within

the native structure, by which proteins with a high proportion of local contacts tend to fold more quickly than those proteins that have more non-local interactions [77].

Detailed studies have also shown that an unfolded polypeptide chain does not follow a single pathway for folding and that many pathways may co-exist, although different paths typically result in the same native state. The different pathways proceed via diverse conformational intermediates, with large energy barriers between the intermediates on the different pathways. The transition state ensemble represents the collection of conformations halfway between the unfolded and native states. Characterisation of transition state ensembles via protein engineering approaches have shown that the ensembles contain conformations with both native and non-native contacts [69, 78, 79].

For many small proteins (typically <100 a.a.), the process of folding occurs in a highly co-operative two-state manner, in which the protein populates either an unfolded or a folded conformation at any point in time. For the majority of proteins, however, the process of folding is often influenced and assisted by the formation of stable folding intermediates. The involvement of intermediates is not limited to large proteins, but has also been identified for some small single-domain proteins that are predicted to have uncomplicated folding profiles [80]. These intermediate states occupy local minima and their structural properties can be characterised as a 'collapsed state', in which a compact, heterogeneous ensemble without well-defined secondary or tertiary structure is populated. In addition, the formation of intermediates possessing 'molten-globule' characteristics, *i.e.* some secondary but little or no tertiary structure, can both assist in the formation of the native state, as these intermediates can possess many characteristics of native-like structure, but can also re-organise, forming kinetically-trapped species (*e.g.* oligomers) that directs folding towards the formation of intermolecular contacts, promoting the formation of amorphous aggregates, or in some cases, highly ordered, thermodynamically-stable amyloid fibres (Figure 1.3). In this description, the processes of folding and misfolding are competing pathways and understanding both the process of folding and the relationship between folding and misfolding represents a fundamental question in contemporary biology.

1.2.2 Protein folding in the cellular environment

Our current understanding of protein folding has resulted largely from the study of isolated proteins *in vitro*, in which the entire polypeptide sequence is available for folding (or unfolding). Within living systems, however, protein folding occurs under a set of very different conditions.

1.2.2.1 The nature of the cellular environment

In contrast to *in vitro* folding, where the entire sequence is available in the cellular environment, the nascent chain emerges from the ribosome at the rate at which translation occurs, which is on the order of 100 peptide bonds per minute in *E. coli* [81, 82]. Moreover, as the nascent polypeptide chain emerges from the ribosome and into the cellular milieu, it encounters a complex environment with a high concentration of macromolecules (up to $400\,g{\cdot}l^{-1}$ [83]) [62]. These macromolecules include auxiliary factors such as molecular chaperones, that assist in folding (*vide infra*) as well as those that are associated with translocation, signalling and post-translational modifications. The high concentration of macromolecules limits the space that would otherwise be available to polypeptide chain for conformational sampling. Both simulations [84, 85] and experiments have shown that the effect of crowding is to shift the equilibrium of protein folding towards more compact structures [86, 87] and also to accelerate both the rates of folding and misfolding [88]. Computational studies have also suggested that the entropy of the nascent polypeptide chain is altered as a result of macromolecular crowding, which alters the thermodynamic and kinetic properties of protein folding by destabilising the disordered form [89].

In addit ion, the nature of efficient protein translation within cells results in formation of polysomes, which act to increase the local crowding effect in the vicinity of the nascent chains. Polysomes are higher order assemblies in which many ribosomes cluster on a single mRNA transcript, creating a concentrated population of nascent polypeptide chains. Cryo-EM studies [90] of polysomes have shown that the ribosomes are oriented on the mRNA in such a way that the nascent chains emerge in at different directions, which is thought to

minimise contacts between neighbouring chains and therefore prevent aggregation.

1.2.2.2 Interaction with molecular chaperones during f olding

Within the cell, up to 40 % of proteins are predicted to interact with molecular chaperones [91, 92], which assist in improving the efficiency and success of folding. Within prokaryotes, a ribosome-associated molecular chaperone, trigger factor (TF), is the first chaperone that interacts with the nascent chain.

TF is 48 kDa multi-domain protein which has dual roles as both a chaperone and a peptidyl-prolyl cis/trans isomerase (PPIase) [93]. TF is known to cycle on and off untranslating ribosomes with a relatively weak affinity (K_d ~1 μM), but this affinity increases in the presence of an emerging nascent chain (K_d=50 nM). This suggests that the cycling of TF on and off the ribosome allows it to inspect the ribosome for the presence of a nascent-chain.

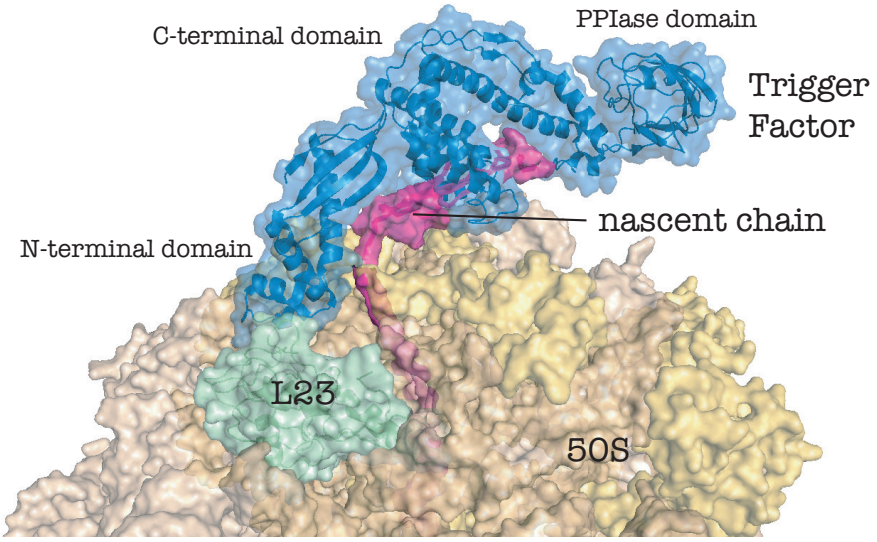

Figure 1.4: A model of a TF-ribosome complex. Modelled onto the crystal structure of the 50S subunit (yellow, PDB 2AW4) is TF (blue, PDB 1W26 [94]), which binds to the ribosomal protein L23 (green). TF forms an arch at the base of the ribosomal exit tunnel which can be occupied by an emerging nascent chain (model of ddFLN$_{646\text{-}838}$, magenta). TF-ribosome complex were modelled by alignment via ribosomal protein L23 in the Cryo-EM structure of the TF-L23 complex (PDB 2VRH [95]), the nascent chain was modelled by Luke Goodsell, UCL.

The X-ray crystal structure of TF from *E. coli* shows that TF has an elongated 'dragon-like' structure formed by three domains: the N-terminal ribosome-binding domain, the

C-terminal chaperone domain and the PPIase domain. The N-terminal domain contains the motif 'GFRXGXXP', which is essential for interaction between TF and the ribosomal protein L23 (Figure 1.4). X-ray and cryo-EM structural studies of the TF-ribosome complex show that TF forms an arch-like conformation over the opening of the exit tunnel [93, 94]. The positioning of TF suggests that it might chaperone the NC by shielding it from the cellular environment as it first emerges from the exit tunnel. Cross-linking experiments have also shown that the entire interior surface of the C-terminal domain can form interactions with the NC [96]. Together with structural studies that show that the interior surface is very hydrophobic, this implies that TF protects the hydrophobic stretches of the NC during the process of co-translational folding, until they became buried within the folded protein structure. TF is also likely to influence the conformation of the NC as it emerges from the ribosome and guides it towards the formation of favourable intermediate states in order to achieve its native fold.

1.2.3 Co-translational protein folding

Aside from the cellular environment and the presence of molecular chaperones, another significant difference that distinguishes protein folding *in vitro* from that which occurs *in vivo* is that within living systems, the polypeptide chain is synthesised on ribosomes, and as the N-terminus begins to emerge from the ribosomal exit tunnel, it has the opportunity to begin sampling conformations before the rest of the protein sequence has been synthesised [97].

Since Anfinsen's early protein folding experiments, many studies have speculated on the events that occur during *in vivo* protein folding, and indeed if the process of folding can begin during synthesis, a process known as 'co-translational folding'. In addition, it has been suggested that an extended, unfolded conformation for the growing polypeptide chain is energetically unfavourable [98], which supports a role for co-translational folding, as do the effects of macromolecular crowding, which favour shifts in the conformational equilibria toward the formation of compact structure [62]. While it is possible that proteins can only

acquire their complete, biologically active structure after release from the ribosome, it is believed that almost all proteins populate folding intermediates co-translationally [98–102].

A number of biochemical and biophysical studies using various techniques have shown that emerging nascent chains adopt native-like structures during co-translational folding, as illustrated in Figure 1.5. Ribosome-bound nascent chains can form disulfide bonds [103, 104] and acquire enzymatic activity [105–107]. The structural properties of ribosome-bound nascent chains can also be recognised by highly specific conformational antibodies [108, 109], probed by limited proteolysis experiments [81, 110, 111] and examined by fluorescence strategies such as FRET [112] and fluorescence anisotropy decay [113, 114], the latter of which has provided an insight into the dynamic behaviour of the nascent chain.

A classical model system used in studies of co-translational folding is firefly luciferase [105, 111, 115–117], which is a 60-kDa multi-domain protein with a mix of α-helical and β-sheet content within its structure. Limited proteolysis studies of luciferase have demonstrated that the first 190 N-terminal residues can form a protease-resistant native-like domain during the process of co-translational folding. This domain, however, is not formed during the renaturation of the isolated protein *in vitro* even in the presence of chaperones [111]. This was a pivotal study which not only demonstrated that co-translational folding was possible, but that folding as it occurs on the ribosome may be different to that which occurs *in vitro* for an isolated protein. Thus the nature of the vectorial emergence of the nascent chain plays a significant role in the way in which the protein forms its native structure.

More recently, the protein SufI, derived from *E. coli*, has emerged as another model system in studies of co-translational folding. Limited proteolysis experiments have shown that emerging nascent chains of this 50-kDa multi-domain protein can form protease-resistant domains, indicating that the protein can fold co-translationally. As with luciferase, *in vitro* renaturation experiments of SufI indicate that it refolds inefficiently as a complete polypeptide and is prone to aggregation [118].

Inspection of the DNA sequence of SufI found clusters of rare codons, corresponding to a low abundance of tRNA [82], which are typically associated with translational pausing. The

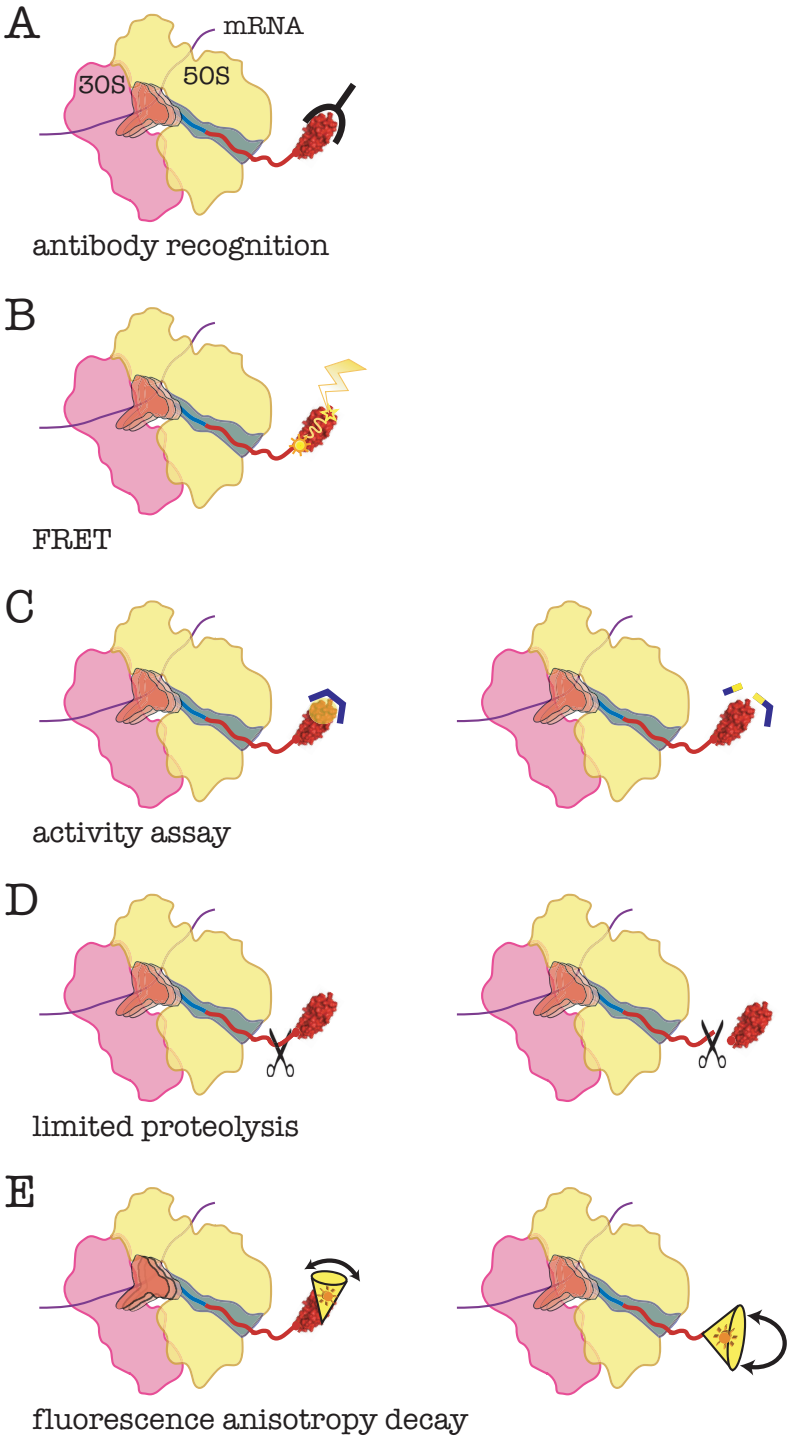

Figure 1.5: Evidence showing protein folding occurs co-translationally, as provided by different techniques, including A: conformation-specific antibody recognition [108, 109], B: FRET [112], C: activity assay [105–107, 115], D: limited proteolysis [81, 110, 111] and E: fluorescence anisotropy decay [113, 114].

locations of these codon clusters were found to be *ca.* 20 residues downstream of each of the domains that were identified by limited proteolysis. Given that the ribosomal exit tunnel can accommodate 20-40 residues [52, 58], the implication is that the location of the rare codons could influence the folding of the SufI protein. Indeed, mutation of the rare codons to more frequently used codons yielded a protein that no longer showed co-translational proteolytic resistance. However, co-translational folding of the protein in the absence of the rare codons could be rescued if the translation reaction was performed at low temperatures [81]. This study strongly suggested that the regulation of translation rate occurs within the cell in a way which actively co-ordinates co-translational folding [81].

1.2.3.1 Structural evidence for co-translational folding

Apart from the host of elegant biochemical and biophysical studies of co-translational folding, recent cryo-EM studies of stalled ribosome-nascent chain complexes have revealed details of the nascent polypeptide chain within the ribosomal exit tunnel, suggesting that it remains largely extended throughout the length of the tunnel [25]. The exact conformation is likely to depend in part on the sequence of the nascent chain, with cryoEM structures of specially-engineered nascent chains [52] giving rise to electron density consistent with partial helical structure. This is also supported by FRET studies of transmembrane domains, which demonstrate that the nascent chain can adopt an α-helical structure [112].

However, the adoption of persistent tertiary structure in the tunnel is probably prevented by its limited dimensions (up 10-20 Å wide [3, 119]), which narrowly confine the nascent chain [25, 51, 52]. Experimental studies have shown that the ribosomal proteins L4, L22, L23, L29 and L32 can make contacts with the nascent chain (Figure 1.6 [52, 120]), suggesting that the ribosomal exit tunnel can sense the nascent chain and potentially influence its movements. Despite the restriction within most of the tunnel, approximately 80 Å from the PTC is the 'exit vestibule', which at 20 Å in diameter, represents the widest part of the tunnel for the nascent chain to sample conformational space before it emerges fully into the cellular millieu [119]. Biophysical studies using cysteine-modifying

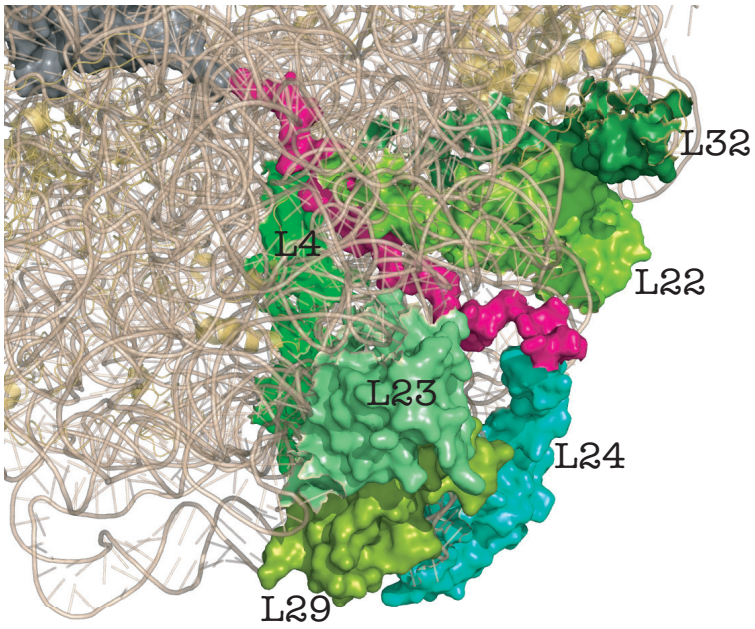

Figure 1.6: Magnified view of the exit tunnel, highlighting the ribosomal protein components (green) and the emerging nascent chain (magenta).

experiments have indicated that simple structural motifs such as α-helices and β-hairpins can be formed within the vestibule [20]. In addition, a number of computational studies have suggested that persistent tertiary structure can occur at a distance of 24-27 residues from the PTC, which corresponds to the exit vestibule [59].

Furthermore, by incorporating experimental data into simulations, theoretical models have emerged that suggest that the role of ribosome within co-translational folding is to reshape the folding landscape [59, 121]. This also depends on the growing nascent chain, whereby the number of amino acids available for folding, and therefore the shape of the energy landscape, change constantly.

There is also a suggestion that the ribosome influences the rate of folding, decreases the folding rate close to the ribosomal surface, which may be a mechanism to prevent misfolding [59]. In addition, it has been found that the presence of the ribosome reduces the number of parallel folding pathways and that more compact, native-like transition-state ensembles were observed [59].

Although some simple structures are predicted to be able to form within the ribosomal exit tunnel, it is clear that most of the folding takes place outside the ribosome. As ribosome-

bound nascent chains are highly dynamic, their structural and dynamic characteristics outside of the tunnel have not been observed by structural techniques such as X-ray crystallography or cryoEM. Elucidating the molecular details of the co-translational folding process would significantly improve our understanding of how protein folding occurs within living systems. Solution NMR is the most promising technique for characterising the nascent chain as it emerges from the ribosome.

1.3 NMR studies of ribosome and ribosome-nascent chain complexes

NMR spectroscopy is a uniquely powerful technique, providing both structural and dynamic information on proteins at atomic resolution.

1.3.1 NMR of the 70S ribosome

As mentioned in Section 1.1.2, NMR spectroscopy has been applied to the study of 70S ribosomes, where the ^{15}N HSQC spectrum of 70S ribosomes was found to contain a collection of resonances that arise almost exclusively from the L7/L12 stalk. The observation of the ribosomal protein L7/L12 is the result of the long, flexible linker that tethers the CTD to the ribosomal body. This allows the L7/L12 CTD to move almost independently of the ribosome, resulting in relatively narrow NMR signals despite its attachment to the large ribosomal body. The CTD domain is therefore capable of rapid tumbling and this is supported by ^{15}N relaxation studies, which have shown that the ribosome-bound L7/L12 CTD has a rotational correlation time of *ca.* 14 ns [34], much shorter than that of the ribosome (*ca.* 2500 ns [122]).

Comparison of the chemical shifts of ribosome-bound L7/L12 compared with those obtained from isolated L7/L12 reveals close similarities between the two species despite their different physical environments [34]. While this strongly indicates that L7/L12 can acquire the equivalent structure on the ribosome as in isolation, its dynamic behaviour on the

ribosome remains to be explored in detail. Moreover, detailed structural information in the form of its orientation and spatial positioning on the ribosome is currently unavailable, and the development of NMR strategies that seek to describe these characteristics is described in Chapter 2.

1.3.2 NMR studies of RNCs

The use of NMR spectroscopy to inspect highly mobile regions on the ribosome, such as L7/L12, provided the motivation for an initial study into the possibility of applying NMR to ribosome-bound nascent chains. This approach was first demonstrated for a pair of tandem-repeat immunoglobulin-like domains (domains 5 and 6) derived from the F-actin cross-linking gelation factor from *Dictyostelium discoideum*, referred to in this study as ddFLN$_{646\text{-}838}$. An RNC was generated for ddFLN$_{646\text{-}838}$ in which the folding incompetent domain 6 formed a flexible linker that tethered domain 5 to the PTC, a scenario reminiscent of ribosome-associated L7/L12.

The use of rapid acquisition methods such as the SOFAST-HMQC experiment [123] on uniformly ^{15}N-labelled RNCs allowed observation of ^{15}N HMQC spectra, which showed that domain 5 adopted a native fold that resembled the structure of the isolated protein. The folded state has since been further studied using ^{1}H-^{13}C correlation spectra, exploiting the enhanced sensitivity afforded by the methyl-containing side-chains [124].

Similar NMR approaches have been applied in the study of barnase-RNCs [125] and also SH3-RNCs using a ^{1}H-^{15}N CRINEPT strategy [126]. The latter revealed that a disordered SH3 domain was observable whilst bound the ribosome, but could acquire its native structure upon release. Together these studies have allowed initial structural and dynamic characterisation of ribosome-bound NCs. However, at present, direct structural information on the nascent chain is limited.

1.3.3 Use of translational diffusion to monitor ribosomal integrity

By virtue of its inherently dynamic nature, the ribosome has a limited lifetime *in vitro*. An important NMR tool to evaluate the integrity of the ribosome is translational diffusion. The use of translational diffusion in this manner follows from extensive studies of the diffusional properties of many isolated proteins [127].

The translational diffusion of a molecule is directly related to its hydrodynamic radius according to the Stokes-Einstein equation:

$$D = \frac{k_{\mathrm{B}}T}{6\pi\eta r} \tag{1.1}$$

where D is the diffusion coefficient calculated from the diffusion experiment, η is the viscosity of the solution, r is the radius of the spherical particle, T is the temperature and k_{B} is Bolzmann's constant.

The hydrodynamic radius calculated using the Stokes-Einstein equation can then be used to estimate the molecular weight according to the following equation:

$$M = \frac{1}{v} \cdot \frac{4}{3}\pi\left(\frac{r_h - \delta}{f_e/f_o}\right)^3 \cdot N_A \tag{1.2}$$

where M is the molecular weight, v is the specific volume, r_h is the hydrodynamic radius, δ is the thickness of the hydrated water layer, f_e/f_o is the shape factor and N_{A} is Avogadro's number.

In this calculation, there are two principal assumptions. Firstly, the thickness of the hydrated water layer is assumed to correspond to a single layer of water molecules, but this parameter is difficult to be determined accurately. Secondly, the shape factor is usually assumed to be 1, as for a perfectly spherical globular protein. This shape factor can be calculated according to the shape of the protein, if it is known.

The diffusion constant is measured using a gradient spin echo experiment. In its simplest form, this experiment consists of an excitation pulse followed by two gradients of opposite sign, separated by a delay. The first gradient acts to encode the position of the spins as a

spatially dependent phase factor. This spatially dependent phase is decoded by the second gradient, but this decoding will only be perfect if there is no movement of spins between the gradients. In practice, this condition is violated due to diffusion-mediated molecular motion during the inter-gradient delay. This leads to imperfect refocusing of the spatially dependent phase, and hence a decay in signal intensity. The decay of signal intensity is described by the Stejskal-Tanner equation [128]:

$$I = I_0 \cdot exp\left(-D(\gamma\delta G)^2(\Delta - {}^{\delta}/_{3}\right) \tag{1.3}$$

The observed intensity, I, depends on the initial intensity I_0, the diffusion constant D, the length δ and strength G of the encode/decode gradients, the inter-gradient diffusion delay Δ and the magnetogyric ratio γ of the detected spin.

In this work, diffusion experiments are used mostly to monitor the integrity of the ribosome complex and the attachment of the NC. A fresh, intact ribosome sample shows slow diffusion with a diffusion coefficient of 1.7-2x10^{-11} m^2s^{-1}. Small proteins have diffusion coefficients on the order of 1x10^{-10} m^2s^{-1}. While the NC is attached to the ribosome, the translational diffusion of the NC is the same as the ribosome. But when measuring simple ^{1}H diffusion experiments, the detected signal is dominated by the ribosome as the ribosomal signals swamp those from the NC. To distinguish between the NC and the ribosome, a ^{15}N-edited diffusion experiment is used, which only detects signals from ^{15}N-labelled species (the NC). The analysis of diffusion data is important in the study of ribosomes and RNCs, as the integrity of the ribosome and attachment of the NC is crucial for reliable analysis and interpretation of the NMR data.

1.4 Introduction to residual dipolar couplings (RDCs)

Residual dipolar coupling is an NMR property that can be used to determine protein structure, domain orientation and dynamics. It is a relatively new NMR method compared to the traditional nuclear Overhauser effect (NOE) technique that has long been used for

structure determination, and complements well the NOE method. It provides distance-independent structural information, while NOE interactions are only visible over a short range (5 Å). Although partial structure determination can be performed with only RDC data, they are more often used as restraints to refine existing structures, as the distance-independent structural information is useful in defining angles, e.g. angles between two helices or the relative orientation of two domains of a protein. In this section, the theory of RDCs, how they are measured and how they are exploited for structure determination and refinement are discussed.

1.4.1 The origin of RDCs

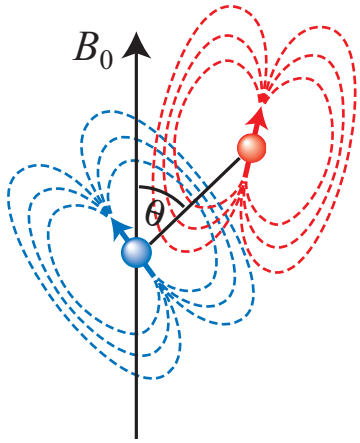

Figure 1.7: Schematic of the dipolar interaction between two magnetic dipoles. θ is the angle between the internuclear vector and the B_0 field.

Dipolar coupling is the mutual interaction between two magnetic dipoles. The size of the coupling depends on the angle (θ) between the internuclear vector and the external applied field (B_0) (Figure 1.7). In solution, molecular tumbling gives rise to rapid modulation of θ. For isotropic tumbling, the dipolar coupling averages to zero as the internuclear vector samples all possible orientations with equal probability. However, this is not the case when the tumbling is anisotropic. In this case, the orientational probability distribution is not spherical, and hence the time-averaged dipolar coupling is non-zero. This non-zero coupling is called the residual dipolar coupling (RDC). The instantaneous dipolar coupling has a magnitude of several kHz, while the degree of induced alignment is usually very small,

on the order of one part in a 1000. Hence, the residual dipolar couplings are typically a few Hz.

The instantaneous dipolar coupling is described by:

$$D_{IS} = \frac{1}{2} D_{IS}^{max} \left(3\cos^2\theta - 1\right) \tag{1.4}$$

where I and S represent the two interacting nuclei.

D_{IS}^{max} is the maximum possible dipolar coupling and corresponds to the internuclear vector lying parallel (or anti-parallel) to the applied field.

The size of D_{max} depends on the magnetogyric ratios (γ) of the two nuclei and the internuclear distance (r), and is given by:

$$D_{IS}^{max} = \frac{-\mu_0 h \gamma_I \gamma_S}{8\pi^3 r_{IS}^3} \tag{1.5}$$

where μ_0 is the permeability of free space.

The expression for the RDC, $\langle D_{IS} \rangle$, can be derived from that for the instantaneous coupling (Equation 1.4) by taking the time-average of the angular term:

$$\langle D_{IS} \rangle = \frac{1}{2} D_{IS}^{max} \langle 3\cos^2\theta - 1 \rangle \tag{1.6}$$

The term $\langle 3\cos^2\theta - 1 \rangle$ depends on both the orientational distribution of the molecule relative to the field and the orientation of the bond vector in the molecular frame, i.e. the structure.

The two contributions can be separated mathematically by first writing $\cos\theta$ as the scalar product between unit vectors corresponding to the internuclear vector in a fixed molecular frame, $\mathbf{r}$, and the (time-dependent) applied field in the same frame, $\mathbf{B}(t)$:

$$\langle \cos\theta \rangle = \langle \mathbf{B}(t) \cdot \mathbf{r} \rangle \tag{1.7}$$

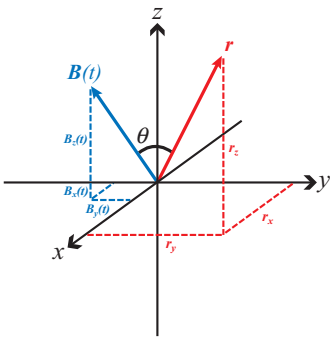

Figure 1.8: The internuclear vector and time-dependent applied field in an arbitrary molecular frame.

The term $\langle \cos^2 \theta \rangle$ may then be written as the matrix product (see Figure 1.8):

$$\langle \cos^2 \theta \rangle = \langle (\mathbf{B}(t) \cdot \mathbf{r})^2 \rangle = \begin{pmatrix} r_x & r_y & r_z \end{pmatrix} \begin{pmatrix} \langle b_x^2(t) \rangle & \langle b_x(t) b_y(t) \rangle & \langle b_x(t) b_z(t) \rangle \\ \langle b_y(t) b_x(t) \rangle & \langle b_y^2(t) \rangle & \langle b_y(t) b_z(t) \rangle \\ \langle b_z(t) b_x(t) \rangle & \langle b_z(t) b_y(t) \rangle & \langle b_z^2(t) \rangle \end{pmatrix} \begin{pmatrix} r_x \\ r_y \\ r_z \end{pmatrix} = \mathbf{r}^T \cdot \mathbf{P} \cdot \mathbf{r} \tag{1.8}$$

The matrix, $\mathbf{P}$,

$$\mathbf{P} = \begin{pmatrix} \langle b_x^2(t) \rangle & \langle b_x(t) b_y(t) \rangle & \langle b_x(t) b_z(t) \rangle \\ \langle b_y(t) b_x(t) \rangle & \langle b_y^2(t) \rangle & \langle b_y(t) b_z(t) \rangle \\ \langle b_z(t) b_x(t) \rangle & \langle b_z(t) b_y(t) \rangle & \langle b_z^2(t) \rangle \end{pmatrix} \tag{1.9}$$

is called the probability matrix, and describes the orientational distribution of the applied field in the molecular frame.

Under isotropic conditions, this orientational distribution, $\mathbf{P}_{\text{iso}}$, is given by:

$$\mathbf{P}_{\text{iso}} = \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix} \tag{1.10}$$

which is equal to $1/3\mathbf{E}$, where $\mathbf{E}$ is the identity matrix.

The anisotropy of the probability distribution is described by the alignment matrix $\mathbf{A}$, which is the traceless part of $\mathbf{P}$:

$$\mathbf{A} = \mathbf{P} - \frac{1}{3}\mathbf{E} \tag{1.11}$$

Now, Equation 1.8 can be rewritten as:

$$\langle \cos^2 \theta \rangle = \mathbf{r}^T \cdot \mathbf{A} \cdot \mathbf{r} + \frac{1}{3} \tag{1.12}$$

and Equation 1.6 can now be expressed as:

$$\langle D_{IS} \rangle = D_{IS}^{max} \cdot \frac{3}{2} \left(\mathbf{r}^T \cdot \mathbf{A} \cdot \mathbf{r} \right) \tag{1.13}$$

In principle, alignment tensor can be expressed in any molecular frame, but to simplify the equations, the principal axis system (PAS) is usually used, as shown in Figure 1.9, which describes the orientation of the alignment tensor.

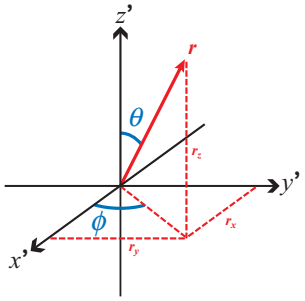

Figure 1.9: The principal axis system. Here, θ and ϕ describes the orientation of the internuclear vector in the PAS.

According to Equation 1.11, the alignment matrix can be written as:

$$\mathbf{A} = \begin{pmatrix} \langle b_x^2(t) \rangle & \langle b_x(t) b_y(t) \rangle & \langle b_x(t) b_z(t) \rangle \\ \langle b_y(t) b_x(t) \rangle & \langle b_y^2(t) \rangle & \langle b_y(t) b_z(t) \rangle \\ \langle b_z(t) b_x(t) \rangle & \langle b_z(t) b_y(t) \rangle & \langle b_z^2(t) \rangle \end{pmatrix} - \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix} = \begin{pmatrix} \langle b_x^2(t) \rangle - \frac{1}{3} & \langle b_x(t) b_y(t) \rangle & \langle b_x(t) b_z(t) \rangle \\ \langle b_y(t) b_x(t) \rangle & \langle b_y^2(t) \rangle - \frac{1}{3} & \langle b_y(t) b_z(t) \rangle \\ \langle b_z(t) b_x(t) \rangle & \langle b_z(t) b_y(t) \rangle & \langle b_z^2(t) \rangle - \frac{1}{3} \end{pmatrix} \tag{1.14}$$

In the principal axis system, the off-diagonal terms of the alignment matrix are all equal to zero, and therefore the alignment matrix can be simplified as:

$$\mathbf{A} = \begin{pmatrix} A_{xx} & 0 & 0 \\ 0 & A_{yy} & 0 \\ 0 & 0 & A_{zz} \end{pmatrix} \tag{1.15}$$

where $A_{xx} = \langle b_x^2(t)\rangle - \frac{1}{3}$, $A_{yy} = \langle b_y^2(t)\rangle - \frac{1}{3}$ and $A_{zz} = \langle b_z^2(t)\rangle - \frac{1}{3}$ are the principal values of the alignment tensor.

Now, the dipolar coupling constant can be expressed as:

$$\langle D_{IS}\rangle = D_{IS}^{max} \cdot \frac{3}{2}\left(A_{xx}r_x^2 + A_{yy}r_y^2 + A_{zz}r_z^2\right) \tag{1.16}$$

To deduce structural information, i.e. the orientation of the internuclear vector **r** in the molecular frame, the alignment tensor needs to be known. The alignment tensor can be derived by fitting a set of measured RDCs to a template structure using a numerical method called singular value decomposition (SVD).

In spherical-polar coordinates, Equation 1.16 can be re-written as:

$$D = D_{\mathrm{A}}\left(\left(3\cos^2\theta - 1\right) + \left(\frac{3}{2}R\sin^2\theta\cos 2\phi\right)\right) \tag{1.17}$$

where D_{A} is the axial component of the alignment tensor scaled for the I-S coupling, which can be expressed by:

$$D_{\mathrm{A}} = \frac{1}{2}D^{max} \cdot \frac{3}{2}A_{zz} \tag{1.18}$$

and R is the rhombicity of the alignment tensor, which is dimensionless and can be expressed by:

$$R = \frac{3}{2}\frac{A_{xx} - A_{yy}}{A_{zz}} \tag{1.19}$$

1.4.2 Degeneracy in RDCs

The available orientations of an intermolecular vector for a single measured RDC shows a strong angular degeneracy relative to the alignment tensor in the molecular frame. In other words, for a single vector (*i.e.* a single type of RDC), measured in one alignment medium, there are many possible orientations in the PAS that all satisfy the measured RDC. These possible orientations lie on the cone-shaped surface, described by Equation 1.17 (Figure 1.10 A).

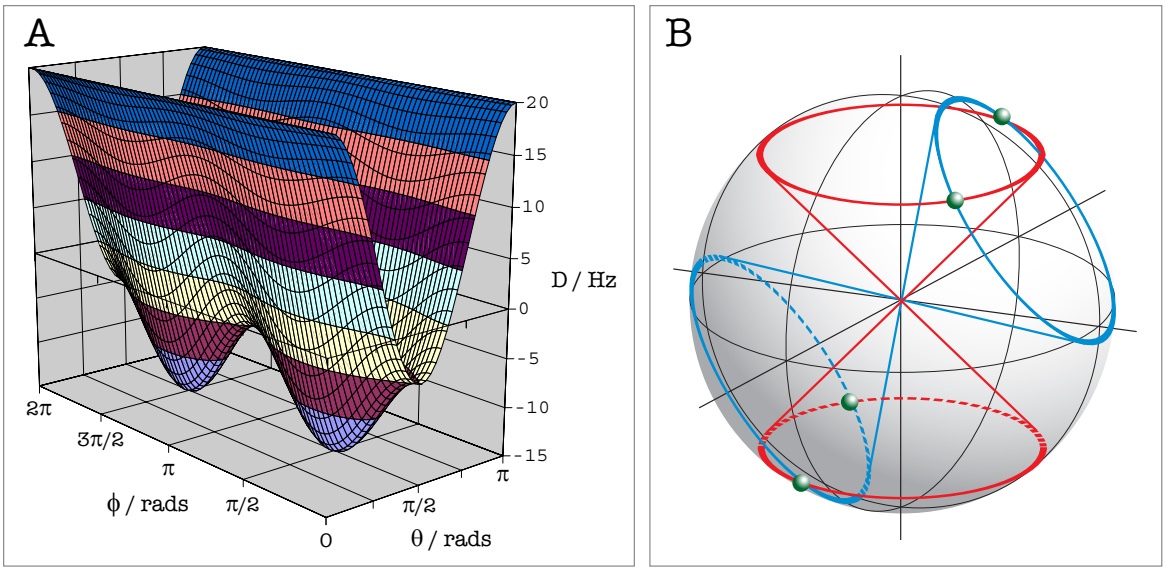

Figure 1.10: **A:** Angular dependence of RDCs in the principal axis system of the alignment tensor showing continuum of polar angles consistent with each value of the measured RDC (simulated for D_a=10 Hz and R=0.33). **B:** Partial raising of degeneracy for the orientation of an internuclear vector whose RDC is measured in two alignment media (axially symmetric alignment tensors).

There are two approaches to raise the degeneracy. The first approach involves measuring two or more RDCs in a fragment of fixed local structure. In this way, the fixed relative orientations between the multiple vectors limits the possible orientations of each vector on their individual cones to those orientations that satisfy both the RDC (on the cone surface) and the relative angles between the vectors. The orientational degeneracy of a planar fragment relative to the alignment tensor is reduced to 8-fold, while the orientation of a chiral fragment is restricted to only 4 possibilities.

The second approach is to measure a single RDC in multiple alignment media. In this approach, the PAS for the alignment tensor in one of the media is considered as the reference molecular frame, and the other alignment tensors then appear tilted in this frame relative to the reference alignment tensor. In a similar manner to the first approach, the possible orientations of the internuclear vector is limited by the requirement that it must lie simultaneously on all the cones that describe the orientation of the vector relative to each alignment tensor. In other words, the vector must lie at the intersections of the cones in the reference molecular frame (Figure 1.10 B). In this case, the degeneracy is reduced to 8-fold

with two alignment media (8 intersections of the two cones) and 2-fold with three or more alignment media.

Of course, these two approaches can be combined by measuring multiple RDCs in multiple alignment media. A combination of a planar fragment and two media yields 2-fold degeneracy and this can be reduced to an unambiguously defined relative orientation of the fragment to the alignment tensor by measuring multiple RDCs from a chiral fragment in two or more media.

1.4.3 Alignment methods

Originally, alignment of a given protein within the external magnetic field was achieved by the intrinsic magnetic anisotropy of the molecular system, for example, for DNA and metallo-proteins [129]. However, for most of the molecules of interest, the alignment induced by the intrinsic magnetic anisotropy is too small to be useful, and therefore alignment media were developed to induce molecular alignment.

There are now several types of alignment media that have been used to achieve partial alignment of biological molecules for RDC measurements, including phospholipid bicelles [130, 131], filamentous phage [132, 133], *n*-alkyl-poly(ethyleneglycol)/*n*-alkyl alcohol mixtures [134], purple membrane fragments [135, 136], acrylamide gel and paramagnetic metal ions (lanthanide binding tag) (Figure 1.11). Phospholipid bicelles, filamentous phage, *n*-alkyl-poly(ethyleneglycol)/*n*-alkyl alcohol mixtures and purple membrane fragments are all liquid-crystalline media. Acrylamide gels are used in both compressed and stretched forms, which result in RDCs with opposite signs [137]. Among these liquid-crystalline media, bacteriophage and purple membranes induce alignment of biomolecules via both electrostatic and steric interactions, whereas the other media only orient molecules by steric interactions alone.

Paramagnetic metal ions with anisotropic magnetic susceptibilities are able to induce alignment of proteins to which they are bound, such as the iron-containing proteins myoglobin and the cytochromes [129, 140–142]. For proteins lacking native metal

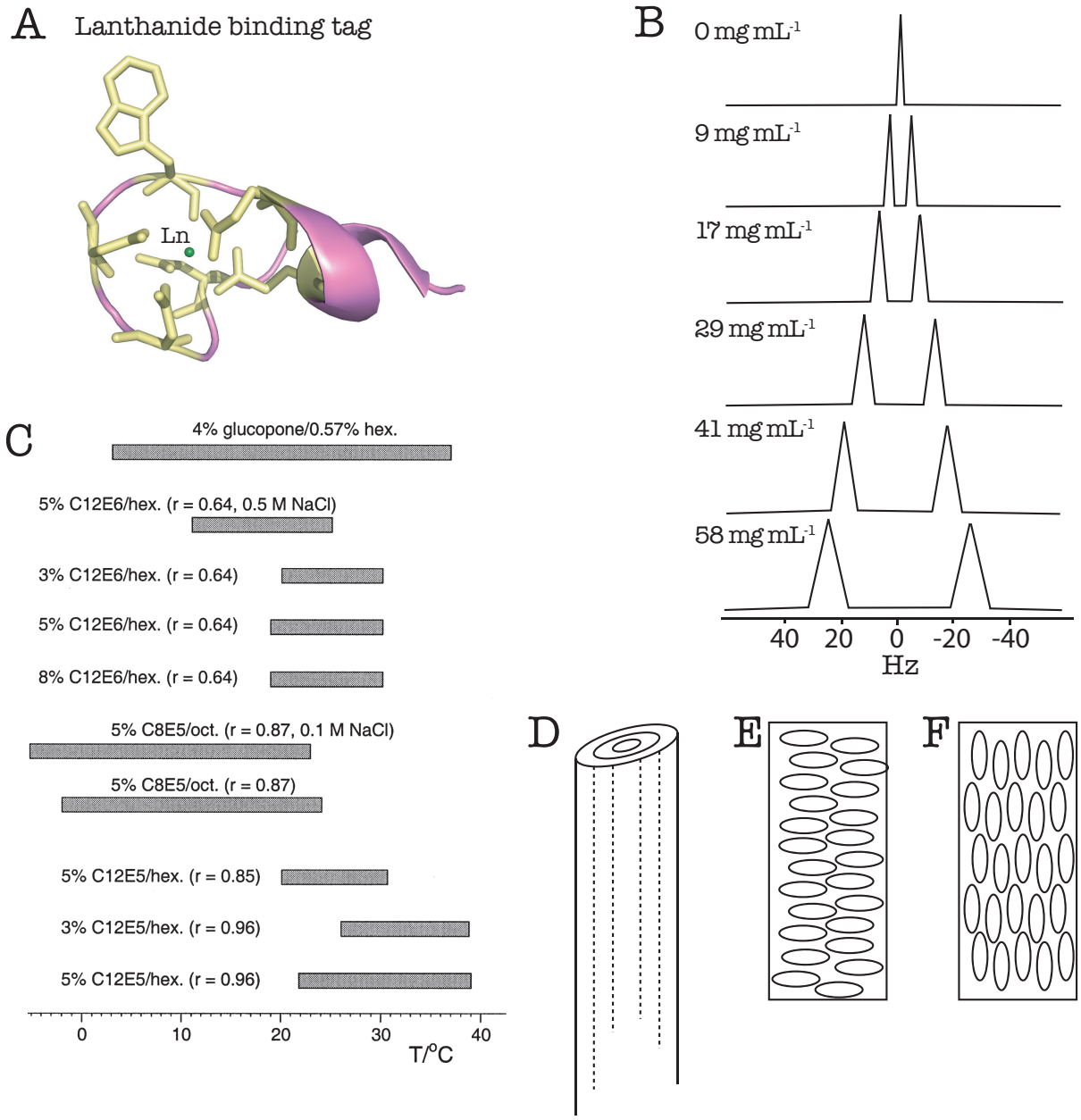

Figure 1.11: A: Lanthanide binding tag (PDB ID: 1TJB). The residues interacting with the lanthanide ion (green sphere) are shown in yellow [139]. B: Schematic diagram illustrates quadrupolar splittings of the ^{2}H NMR signal from different concentrations of Pf1 in 90 %/10 % H_2O/D_2O at 25 °C (produced according to [132]). C: Temperature ranges of stable lamellar phases composed of C_mE_n/alcohol/H_2O mixtures (adapted from [134]). D: Alkyl polyethylene glycol/alcohol mixtures form bilayers that align parallel to the magnetic field [134]. E: Oblate cavities formed in a stretched polyacrylamide gel. F: Prolate cavities formed in a compressed polyacrylamide gel.

binding sites, synthetic metal binding motifs can be incorporated at the DNA level. The lanthanide binding tag (LBT) (Figure 1.11 A) is a short peptide sequence derived from modifying the EF-hand motif of a calcium-binding protein [139]. The sequence **YIDTNNDGWYEGDELLA** binds trivalent lanthanide with high affinity [143] as determined using fluorescence spectroscopy [144], and has been shown to have minimum effect on the protein structure [143]. The dissociation constant between the peptide and trivalent lanthanide ion depends on the ionic radius, where the smallest ion Lu^{3+}, has K_d=130 nM and the largest La^{3+} has K_d=4 μM [143]. The isotropic reference state is obtained by using diamagnetic ions with similar ionic radii (such as La^{3+}, Lu^{3+} and Y^{3+}) to the paramagnetic ions. In this way any structural changes between the isotropic and aligned states are minimised [145]. For recombinant proteins with LBTs, the binding affinity might change from that for the tag alone, and the K_d between protein-LBT and the ion is often evaluated using fluorescent spectroscopy. This is important as presence of excess paramagnetic ions in the NMR sample causes line broadening through enhancing relaxation [146]. The presence of a LBT and capability to bind lanthanide ions also provides the opportunity to obtain structural information from pseudocontact shifts (PCS) and Föster resonance energy transfer (FRET) [144, 145, 147].

Filamentous bacteriophage is probably the most widely used liquid crystalline medium. The type of filamentous phage commonly used for RDCs experiments is Pf1, from *Pseudomonas aeruginosa*, which forms long rod shape of 1960 nm in length and 6.7 nm in diameter. There are approximately 7000 copies of the helical coat protein arranged symmetrically on the surface [148], which together with the shape, give the phage particles a large anisotropic magnetic susceptibility. This means they are able to align strongly in a magnetic field and hence induce anisotropic tumbling [149]. Phage particles are negatively charged at physiological pH [150], and are thus suitable for aligning negatively charged proteins and DNA/RNA molecules [151], but might interact too strongly with biological molecules that are positively charged. Magnetic alignment of Pf1 phage can be measured from the 1D ^{2}H NMR spectrum, which there show a splitting of the HOD deuterium

signal caused by the incomplete averaging of the deuterium quadrupolar interaction for water molecules bound to the surface of the phage particle. The size of the splitting is approximately proportional to the phage concentration between 10-60 mg ml^{-1} (Figure 1.11 B), and the alignment is stable over the temperature range 5-35 °C once alignment has been achieved [132, 150].

The *n*-alkyl-poly(ethyleneglycol)/*n*-alkyl alcohol mixtures described by Rückert and Otting [134] represent another commonly used liquid crystalline medium for partial alignment of macromolecules. There are different alkyl-(polyelethyleneglyco)/*n*-alkyl alcohol mixtures available for applications at different temperatures (Figure 1.11 C)[134]. The different polyethylene glycol components of the mixture are termed C*m*E*n* (*m* is the number of carbon atoms in the linear alkyl chain and *n* is the number of glycol units [134]. The *n*-alkyl-poly(ethyleneglycol) is the component that forms the liquid crystal while addition of the *n*-alkyl alcohol promotes formation of the liquid crystal phase at lower concentrations of *n*-alkyl-poly(ethyleneglycol). The advantage of this medium is that it can be used under a wide range of conditions. It is relatively insensitive to pH and salt, and can be used from 0 °C to about 40 °C. Importantly, it also has limited binding affinity for macromolecules.

The other liquid crystal media are only usable under a more limited range of conditions. Purple membranes are naturally-occuring two-dimensional crystals formed by the membrane protein bacteriorhodopsin and the surrounding lipids in the cell membrane of the archaeon *Halobacterium salinarium* [135, 137]. In the magnet, disc-like purple membrane fragments are aligned perpendicular to the main magnetic field. This medium tends to interact with proteins strongly [135, 136], which gives rise to line broadening and and alignment tensor that is not solely dependent on the structure of the molecule. DMPC/DHPC [130] mixtures are susceptible to degradation by hydrolysis at low pH [152], while DIODPC/CHAPSO mixtures only form stable liquid crystalline phase at lower than physiological pH [153], and are therefore limited in their application to studing biological macromolecules. Cetylpyridinium chloride/hexanol mixtures are only stable in high-salt environments [154], and the stable phase can be disrupted by high protein concentrations. Another member of this family,

cetylpyridinium bromide/hexanol, is more stable in lower ionic strength conditions [155].

1.4.4 Methods to measure RDCs

The dipolar coupling Hamiltonian has the same form as the scalar coupling Hamiltonian, which means it is not possible to selectively re-introducing evolution under the dipolar coupling Hamiltonian without also re-introduce evolution under the scalar coupling Hamiltonian. In other words, the dipolar coupling constants can not be measured alone experimentally. However, the pure scalar coupling (J) can be measured under isotropic conditions, in which the dipolar couplings (D) all average to zero. Measurement under aligned conditions yields splittings that are then the sum of the scalar and dipolar couplings. The dipolar coupling is then calculated by taking the difference between the isotropic splittings (J coupling only) and the partially aligned splittings (J+D).

The experimental approaches for measuring scalar coupling constants are well-established, and the same approaches can be used for measuring RDCs. There are two principal methods for measuring coupling constants, the *frequency-based* approach and the *intensity-based* approach. The frequency-based approach is relatively more straightforward, with the splitting being extracted by singly measuring the chemical shift differences of the multiplet components from the processed spectrum. The intensity-based approach requires encoding the coupling in the intensity of the observed signal. In the most common J-modulation approach [156], this is achieved by using variable coupling evolution over some delay to modulate the relative proportion of transferable magnetisation at some point in the pulse sequence, so that the intensity of the observed signal depends on the length of the delay and the size of the coupling. The experiment is repeated with different lengths of delay to obtain modulation of the observed intensity, which is then fitted to extract the coupling constant. An alternative approach is called phase-modulated J-spectroscopy, where both magnetisation components of J-interconverted spins are transferred to observable signals in two separate sub-spectra. The coupling can then be extracted from the signal intensity ratio between the two spectra. In general, the intensity-based approach yields more precise measurements of

the coupling than the frequency-based approach. However, the intensity-based approach is more susceptible to systematic errors, which can arise from differential relaxation, passive couplings or cross-correlation.

The type of approach for measuring dipolar couplings also partly depends on the type of couplings to measure. For proteins, RDCs are most commonly measured for the backbone, because the dynamics of the side-chains introduce complications in the alignment tensor calculation. Several backbone RDCs in the peptide plane can be measured, including N-H, N-C', H^N-C', C'-C^α, N-C^α and H^N-C^α couplings. The other backbone RDCs that can be measured are at the junction of the peptide plane, *i.e.* the C^α-H^α and C^α-C^β couplings. Proton-proton couplings are less often measured routinely, because the measurement is complicated by the presence of multiple significant couplings to each proton. Most applications of RDCs for protein structural analysis focus on one-bond proton-heteronuclear couplings. These couplings are large compared to the signal linewidth, which ensures the accuracy of the measured splittings from the corresponding multiplets. From a structural perspective, these one-bond H-X or X-X couplings are easier to interpret than proton-proton couplings, as the internuclear distance is defined by the bond-length, and therefore the observed RDCs depend only on the angles of the internuclear vectors. In addition, for some two-bond couplings between nuclei in the protein backbone, the internuclear distance is fixed by the geometry of the peptide plane, whereas for proton-proton couplings, the internuclear distance may not be well-defined by the local structure.

Although the HX and XX couplings can be obtained from simple 'undecoupled' spectra simply by removing the decoupling pulse from the standard pulse sequence, additional considerations are required. For example, in a normal ^{1}H-^{15}N heteronuclear single quantum correlation (HSQC) experiment, each residue gives rise to one cross peak. To be able to measure the N-H splitting, the decoupling in the ^{15}N dimension is removed and each resonance thus appears as a doublet, resulting in twice the number of peaks as in the decoupled spectrum, leading to more overlap of resonances and difficulties in assignment and measurement. To circumvent such problems, the 'undecoupled' spectra are often recorded

in a spin-state-selective mode. This means that the two components of the relevant doublet are separated into two sub-spectra, where the number of peaks in each sub-spectrum is the same as in the standard decoupled spectrum. In addition, this approach allows accurate measurement of peak positions for doublets where the linewidth is comparable to the size of the splittings. In this case, the measured peak positions in an unseparated spectrum would give rise to underestimated splittings. The in-phase/anti-phase (IPAP) approach is the most common technique used for measuring J coupling and dipolar coupling from simplified two-dimensional NMR spectra [157]. It records two spectra, in which the modulation of the coupling during the relevant indirect shift evolution time period is arranged to have either a sine or a cosine dependence. The resulting sub-spectra then have the doublet components either in-phase or anti-phase. Addition and subtraction of the in-phase and anti-phase sub-spectra yields two different sub-spectra that contain only the upfield or the downfield component. The drawback of the IPAP approach is that the insertion of a preparation period of length 1/(2J) into the pulse sequence gives rise to loss of sensitivity relative to the standard sequence. Additionally, the separation of the doublet components can sometimes be imperfect due to J mismatch and differential relaxation effects. The problem of J mismatch is unavoidable when a range of coupling constants are present. This is generally not a concern for spectra recorded on isotropic samples, where the one- and two-bond HX and XX J couplings are usually relatively homogeneous. However, this problem will be noticeable in aligned spectra, where the J+D values can vary significantly according to the size and sign of the D values.

1.4.5 Structural information from RDCs

1.4.5.1 RDCs for folded proteins

In favourable cases, RDCs can be used to determine protein backbone folds directly.

In the MECCANO approach [158], for RDCs in the peptide plane (H-N, H-C', N-C' and C'-C^{α} and two at the chiral plane junctions (C^{α}-H^{α} and C^{α}-C^{β}) are measured in two alignment media. The structure calculation algorithm uses the peptide plane RDCs to

simultaneously determine the two alignment tensors and orient the peptide planes within the global molecular frame. The orientation of the peptide planes is two-fold degeneracy and any ambiguities are resolved using the RDCs measured at the plane juction.

More commonly, RDCs are used in refinement of existing structures. For example, they can be used to obtain a solution structure by refining the X-ray crystal structure. Furthermore, they can also be used to refine solution structures derived using traditional NOE restraints, since the NOEs provide only short-range distance restraints (< 6 Å), whereas RDCs provide complementary distance-independent orientational restraints.

The orientational dependence of RDCs can also be exploited to determine relative domain orientations within multi-domain proteins. This is slightly different from structural refinement, where the RDC restraints are used to refine the orientations of individual bond vectors to improve the structure of the protein; in determination of relative domain orientations, alignment tensors are calculated for individual domains assuming the domain structure is rigid. In a protein with fixed relative domain orientations, the alignment tensor should be common to all domains, and hence the domain orientations can be adjusted such that the individually calculated alignment tensors coincide.

1.4.5.2 RDCs for unfolded proteins

The recently discovered prevalence of intrinsically disordered proteins, and their importance in a number of disease states, has required significant changes to the traditional structural biology picture, in which proteins adopt well-defined, folded structures with only small-amplitude conformational perturbations [159, 160]. These folded proteins can usually be well represented by a single average structure, with dynamic processes characterised by one or two parameters describing the magnitude and frequency of excursions from the average structure. Nevertheless, improved ensemble-based approaches have recently been developed for describing the structure and dynamics of folded proteins, highlighting that even structures for well-folded proteins are not fixed [161, 162].

For unfolded proteins, the concept of a single average structure becomes meaningless,

with the protein undergoing massive conformational changes on rapid timescales. For an accurate description of the protein's structure and dynamics, it is essential to adopt ensemble-based approaches. However, using NMR-derived experimental data to derive a representative ensemble is more challenging than the traditional approaches in which the NMR data is used for single-structure generation.

One concern of the ensemble approach is the size of the ensemble, which is required to adequately reproduce the experimental data. The problem is ill-posed in that an infinite number of ensembles can satisfy the data. When applied to folded proteins, it is found that only a small ensemble size is required to reach good agreement with the measured NMR data. This is a reflection of the small amplitude of the conformational changes occurring in well-folded proteins. However, for intrinsically disordered proteins, much larger ensemble sizes may be required.

In brief, the ensemble approach as applied to folded proteins involves restraining the ensemble by minimising the difference between the back-calculated and ensemble-averaged NMR parameter and the measured value of the parameter. Application of the approach to unfolded proteins requires more careful generation of the starting ensemble, and increased computational costs associated with evaluation of the ensemble-averaged NMR parameters for large ensembles.

The latter consideration is particularly relevant for RDCs, where the evaluation of the RDCs for each member of the ensemble requires first deriving their distinct alignment tensors. Early interpretations of RDCs measured on intrinsically disordered protein focused on a qualitative comparisons between the measured data values, and those back-calculated from random ensembles based on statistical coil models. Initially, it was thought that the mere presence of non-zero RDCs for supposedly disordered proteins was indicative of residual structure [163]. However, it was later demonstrated theoretically than even random flight chains would show non-zero RDCs [164, 165]. Subsequent refinements of the model to account for the specific conformational preference of polypeptide chains were able to accurately reproduce the RDC profiles of completely disordered proteins [166, 167].

The focus then shifted towards detecting the differences between RDCs predicted from the sequence assuming complete disorder and the measured RDCs that could indicate residual structure or transient long-range contacts along the chain. α-synuclein shows an RDC profile which deviates significantly from that predicted for the random ensemble at the N- and C-termini. Agreement between the simulated and measured RDCs was improved by filtering the random ensemble according to criteria based on the presence of long-range contacts between segments of the polypeptide chain. The filtered ensemble generated by selecting preferentially for conformers with a contact between the N-terminal and C-terminal segments gave a simulated RDC profile that closely matched the measured data [168].

More recently, interpretation of RDC data from intrinsically disordered proteins has been aimed at using the measured data to directly derive the corresponding conformationally-biased ensemble. Two approaches have been developed. In one approach, the RDC data are used to select sub-ensembles from the random ensemble that most closely reproduce the measured RDCs [169]. In the second approach, the RDCs are used to construct a pseudo-potential energy term that is used to drive the evolution of a random starting ensemble in a restrained ensemble MD simulation [170, 171]. Such MD simulations have relied on development of fast algorithms for determination of the alignment tensor, since the alignment tensor must be calculated for every member of the ensemble at every time-step of the simulation. While both approaches have improved the quantitative understanding of RDCs measured on IDPs, there remain some difficulties, such as determining the minimum ensemble size required to describe the data without over-fitting and ensuring that the representative ensemble obeys Boltzmann statistics. Interpretation of RDC data from IDPs remains more challenging than for well-structured proteins.

1.4.5.3 RDC studies of large molecular complexes

The application of backbone RDCs for structure determination and refinement is limited to systems with molecular weights of less and approximately 100 kDa [172, 173]. The recent development of methyl-TROSY-based experiments on highly deuterated protein samples that

contain selectively protonated methyl groups in the isoleucine, leucine and valine residues has lifted the upper limit of NMR observability of large molecular complexes [174–176]. This has enabled application of RDCs to molecules with high molecular weight, as shown by the study of the 360 kDa 'half-proteasome', which demonstrated that methyl ^{1}H-^{13}C RDCs can be measured for very high molecular weight complexes [177].

1.5 Summary

This introduction has summarised the current understanding of ribosome structure and the proposed mechanism of the translation process from the snapshots of ribosome structure captured at different stages of translation. However, studies of the translational product, the nascent polypeptide chain, have not able to provide detailed structural information or mechanistic detail of the process of co-translational folding. NMR spectroscopy has been used in an attempt to extract structural and dynamical information on the nascent polypeptide chain, but as yet, only chemical shift information has been obtainable, which although provides a powerful insights into the folding of nascent chains on the ribosome but not direct structural information. The primary aim of this work is to establish the use of NMR spectroscopy, via measurement of RDCs, to study the structural preferences of RNCs in order to begin to develop three-dimensional structural models. Building upon previous NMR studies of the ribosome, which use the L7/L12 stalk region as a model for the NC, an RDC approach was initially evaluated and developed for the 70S ribosome. This is the focus of Chapter 2. This is then followed by Chapter 3, where the RDC methodology developed for the 70S ribosome was applied to the study of ribosome-nascent chain complexes.

Chapter 2

NMR investigations of the 70S ribosome particle using residual dipolar couplings

2.1 Introduction

The ultimate goal of this work is to use RDCs acquired on nascent chains to probe their co-translational folding as they emerge from the ribosome. Previous NMR studies of the ribosome have identified L7/L12 as a highly mobile and dynamic region (Section 1.1.2). Therefore to examine the feasibility of using RDC to probe ribosome-bound nascent chains, the ribosome-bound L7/L12 is an ideal model system. In particular, generating information about their orientation and other structural and dynamic characteristics, will provide a significant step forward to developing three-dimensional structural models. These flexible units therefore represent a model system upon which to develop the required techniques for measuring RDC data on nascent chains (Figure 2.1). Moreover, the study of the stalk region of the ribosome represents an interesting system in itself, due to its pivotal role in translation. However, elucidating the functional aspects of the L7/L12 protein has been hampered due to its flexibility precluding observation of well-defined electron density in the many recent crystallographic and cryo-EM structures of the ribosome. Therefore, acquisition of RDC data on ribosome-bound L7/L12 provides a means to improve the current understanding of this protein's structure, dynamics and ultimately, its function.

This chapter will therefore explore the development of this approach to study L7/L12, both in isolation and as it is bound to the ribosome.

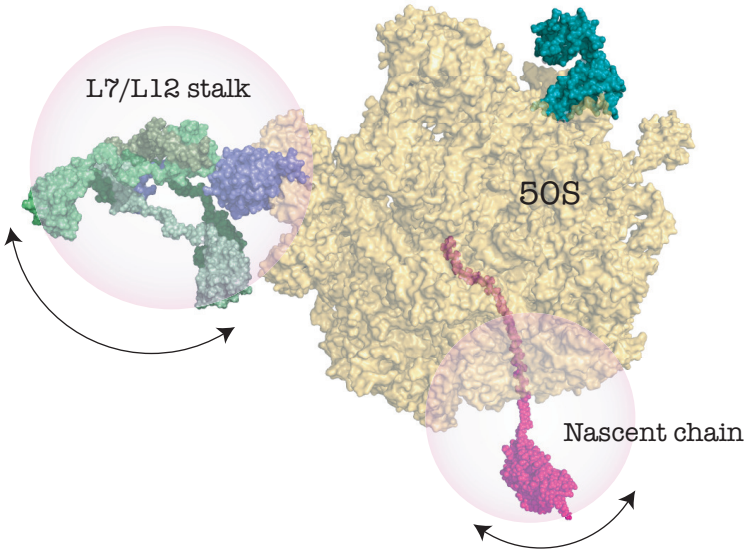

Figure 2.1: Structure of the 50S ribosome (yellow) with models of the L7/L12 stalk region (L7/L12 in green and L10 in violet) and the ddFLN$_{646\text{-}838}$ nascent chain (magenta) illustrating the potential similarities in dynamics of L7/L12 and the nascent chain. The four copies of the L7/L12 protein are shown in different shades of green. For PDB IDs, see Figure 1.1.

2.1.1 The structural and functional characteristics of the L7/L12

2.1.1.1 Characteristics of the ribosome-bound L7/L12

As discussed in Section 1.1.2, the stalk region of the ribosome, and in particular L7/L12, is highly dynamic, and interacts with a number of auxiliary factors, many of which are GTPases, including IF2, EF-Tu, EF-G, and RF3 [178]. These are involved in mediating the translation process as it occurs on the ribosome [2].

The protein L7/L12 is unique among the ribosomal proteins, because it is the only protein present in multiple copies. There are thought to be four copies of L7/L12 in the *E.coli* ribosome [182, 183] and 6 copies in eukaryotes and archeae [184]. The protein L7 and L12 are identical in amino acid sequence, apart from a post-translational modification which results in an acetylated N-terminal serine in L7 [185]. Studies have shown that the ratio between L7 and L12 present on the ribosome varies with the cell growth [182], in which the

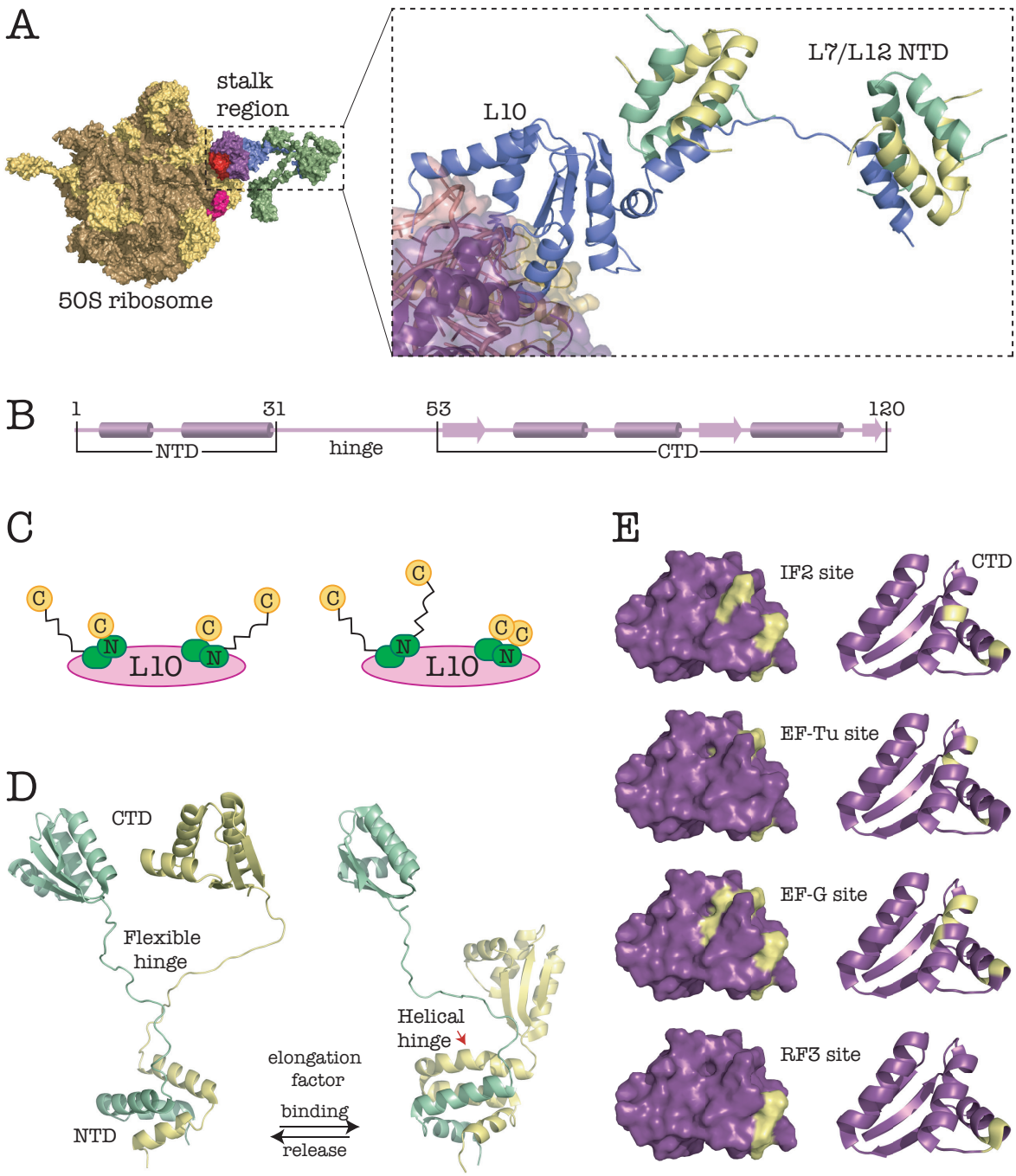

Figure 2.2: **A:** Front view of the 50S ribosome with the stalk region highlighted (as Figure 1.1 A) and a cartoon representation of *E.coli* L10 and the N-terminal domain of two L7/L12 dimers. The L7/L12 and L10 proteins were modelled onto the structure by an alignment with the crystal structure of the *Thermus thermophilus* 50S ribosome, PDB ID 3I8I [23]. **B:** Domain architecture of L7/L12 annotated with secondary structure. **C:** Two possible models of L7/L12 dimers on the ribosome, with either one extended and one retracted CTD within both dimers (LHS) or one dimer with both CTDs extended and one with both CTDs retracted (RHS). **D:** Model of L7/L12 dimer as a molecular switch with the hinge region extended or retracted depending on elongation factor binding, based on X-ray crystallography and NMR data [22, 180]. **E:** Residues in the C-terminus domain of L7/L12 interact with translation factors IF2, EF-Tu, EF-G and RF3 (highlighted in yellow) [181].

L7 form being favoured under restricted growth conditions [186], and the L12 form being the the more active form involved in GTPase-induced GTP hydrolysis [187].

The L7/L12 protein are bound to the ribosome via protein L10, whereby the C-terminus (residues 71-165) of L10 binds multiple copies of L7/L12 and the N-terminus (residues 1-69) binds the $(L7/L12)_4$·L10 pentameric complex to the ribosome [188] via interaction with L11 and L10/L11 binding region of 23S RNA (Figure 1.1 A). Together, this collection of proteins form a large part of the GAR in ribosome.

Although atomic details of much of the GAR have been revealed in the crystal structures of the 50S subunit [2], regions such as the L7/L12 stalk give only fragmented or weak density in cryo-EM [178] and X-ray diffraction maps [21, 189]. Their absence in high-resolution structures probably results from the fact that they do not occupy well-defined positions on the individual ribosomes, which is supported by nanogold labeling experiments, in which the L7/L12 CTD was found to appear at multiple sites on the ribosome [190]. Nevertheless, the CTD of L7/L12 can be observed in a cryo-EM density map when its motion is restricted by the binding of EF-G [191].

Early studies in the late 60s and early 70s had already demonstrated that L7/L12 is required for polypeptide synthesis on the ribosome [192] and that it existed mainly as a dimer in solution [186]. However, it was unknown if all four copies of L7/L12 are required for ribosome function. More recent studies have found that protein synthesis on the ribosome can indeed be supported by the presence of only one CTD in each L7/L12 dimer (80 % activity compared to wide-type) [193] while ribosomes with only one dimer show less than 50 % activity in protein synthesis [194]. These observations suggest therefore that only one CTD is functioning within each dimer at any one time. Aside from the CTD, the hinge region has also been shown to be required for ribosome activity [195], which indicates the importance of the need of flexibility in the CTD for ribosome function, which is highly dependent on the conformation of the hinge region.

2.1.2 Structural and function properties of the L7/L12 protein

Structural investigations of the isolated L7/L12 protein has shown two independent structured domains linked by a flexible hinge [22] (Figure 2.2 B&D). The all-helical N-terminal domain (NTD) (residues 1-31) is the site of dimerisation and ribosome binding (Figure 2.2 A). It was found that the NTD is necessary and sufficient for dimerisation [196]. In addition, NMR studies of the $(L7/L12)_4$·L10 pentameric complex showed that the first 39 residues of L7/L12 are involved in the interaction with L10 [197], which is supported by the X-ray crystal structure of the L10-L7/L12 NTD complex from *Thermotoga maritima* [29]. The globular CTD (residues 53-120) on the other hand, has been shown to be highly dynamic with respect to the NTD and is the site of interaction with the GTPase factors. Translation factors IF2, EF-Tu, EF-G and RF3 bind the ribosome via overlapping regions concentrated between residues 66-82 in the CTD of L7L12 (Figure 2.2 E) [181].

The flexibility of the L7/L12 dimers on the ribosome has been the subject of several studies. One X-ray crystal structure [180] exhibits a helical conformation for the hinge region, which can only be adopted for one hinge region within the dimer at any one time, since there will be a severe steric clash if both hinge regions are helical simultaneously. Subsequent NMR studies on L7/L12 bound to the ribosome suggested that the CTDs of only two out of the four copies of L7/L12 are observed [34, 179]. This observation suggests two possible models for L7/L12 dimers bound to the ribosome (Figure 2.2 C). Either both dimers contain one extended and one retracted CTD, or one dimer has both CTDs extended and the other has both CTDs retracted [179]. However, the latter model is infeasible since two retracted CTDs would induce steric clashes and does not agree with the biological studies discussed above, which have suggested that only one CTD is active in each dimer at any one time. Therefore, the model that each dimer has one extended CTD and one retracted CTD is the most plausible in light of currently available experimental data. This is also supported by analysis of ^{15}N spin relaxation and small-angle X-ray scattering data [198] of the isolated L7/L12 dimer, which has shown an anti-correltion between the positions of the two CTDs relative to the dimeric NTD. Together with the fact that the CTD of L7/L12 interacts with

GTPases, these observations led to the proposal that L7/L12 acts like a molecular switch on the ribosome, whereby each monomer alternates between a helical and extended linker conformation in response to translation factor binding and release (Figure 2.2 D) [22].

Although most studies have suggested that L7/L12 is more than 90 % dimer at micromolar and submicromolar concentrations [186, 199–201], a tetrameric form has also been proposed [202–204]. An early study in 1981 reported equal disassociation constants of 29 μM for monomer/dimer and dimer/tetramer equilibria [204]. However, a more recent study using fluorescence spectroscopy reported a much smaller dimer-to-monomer dissociation constant of 30-300 nM [205], with no tetramer observed even at concentrations above 100 μM. Therefore, while the presence of a dimer-tetramer association process has been demonstrated, the strength of this interaction is not currently well defined.

2.1.3 Experimental approaches to measure RDCs on the ribosome

To understand further the orientation and conformational preferences of L7/L12, an RDC approach was developed as it can yield distance based information on two aspects of L7/L12 on the ribosome. Firstly, they report on the local structure of the protein. Secondly, they define the orientation of the molecular alignment frame relative to the field. In this work, a comparative approach was adopted to study the local structure of the L7/L12 protein, whereby sets of RDCs were measured and analysed for both ribosome-bound and isolated L7/L12.

Since RDCs for a single vector in one alignment medium result in high orientational degeneracy for each vector in the alignment frame, a single set of RDCs in one medium has limited application for structural refinement. Multiple RDCs measured within the peptide plane, and in multiple media would be desirable. For the second approach where the focus is on the orientation of the alignment tensor, the domain structure is considered to be fixed, and hence the degree of degeneracy of the derived alignment tensor in the molecular frame is much less than that associated with a single vector. In fact, measuring a single set of RDCs in multiple media is enough to unambiguously define the orientation of the alignment tensor.

Measurement of RDCs on the ribosome presents several challenges. In particular, the limited concentration of the samples shapes the measurement strategy. The relative accuracy of the measured RDCs is dependent on both the intrinsic magnitude of the RDC and the signal-to-noise ratio and peak linewidths of the experimental data. This means that the choice of the types of RDCs to measure is limited not only to those of intrinsically large magnitude but also to those that can be measured with high sensitivity.

The most commonly measured RDC is the backbone amide N-H RDC, since the size of the coupling is intrinsically large and can be measured with sensitive pulse sequences relatively easily. The H^{α}-C^{α} RDC is the largest one-bond coupling within a protein, with a magnitude approximately twice that of the N-H coupling, but the measurement is technically more demanding.

In this chapter, a detailed investigation of intact ribosomal particles using NMR spectroscopy is described. In particular, the measurement and analysis of residual dipolar couplings on the ribosome is demonstrated for the first time.

2.2 Results and Discussion

2.2.1 Considerations for studying the ribosome by NMR

The challenges associated with NMR investigations of the ribosome arise primarily as a result of its large molecular weight (2.5 MDa) and the limited lifetime of the samples. Indeed, the quality of the sample is crucial for obtaining meaningful NMR data: the samples must be isotopically enriched to ensure that the maximum possible signal can be obtained. This is pertinent due to the limited working concentrations that are achievable for the ribosome, which are at least an order of magnitude lower than typical protein NMR concentrations. Above this threshold (*ca.*10 μM), the increase in viscosity manifests as significant broadening of resonances [206], and the sample lifetime is reduced. In addition, the sample needs to be homogeneous, as impurities arising from co-factors, for example, EF-G, may interact with the L7/L12 protein resulting in spectral changes. The presence of other isotopically-labelled molecules may also result in extra resonances appearing in the spectra, potentially resulting in misinterpretion of the data. The lifetime of the samples is also influenced by the presence of bacteria that can cause rapid degradation of the ribosome. Together, these are of a particular concern to studies of the ribosome as the low working concentrations result in signals that are inherently weak and are very often close to the noise threshold, and therefore even weak signals from contaminating species can often obscure the signals of interest. As a result, the integrity of the ribosome is monitored rigorously throughout the course of NMR data acquisition, which is essential to ensure that the NMR signals are derived exclusively from intact ribosome species.

2.2.1.1 Isolation and purification of intact 70S ribosomes

The preparation of intact 70S ribosome complexes is described in detail in section 5.1.12 on page 192. In brief, an enriched medium composition enables the growth of cultures to high cell densities (section 5.1.3 on page 183) before harvesting. Following lysis, the ribosomes were separated from the other soluble cell components by centrifugation through a 30 %

(w/v) sucrose cushion, after which the ribosomal material is recovered as a pellet at the bottom of the tube. The 70S monosome fraction is then purified by centrifugation through a 10-35 % (w/v) sucrose gradient, as shown in Figure 2.3 A.

The purity of the ribosome samples used in NMR experiments were evaluated by SDS-PAGE, which typically show a collection of bands, each one corresponding to a single ribosomal protein (example shown in Figure 2.3 B, with the bands highlighted in the black box). Within the 70S complex, the largest ribosomal protein is S1(61 kDa), while the other ribosomal proteins have molecular weights ranging between 5.4 and 30 kDa.

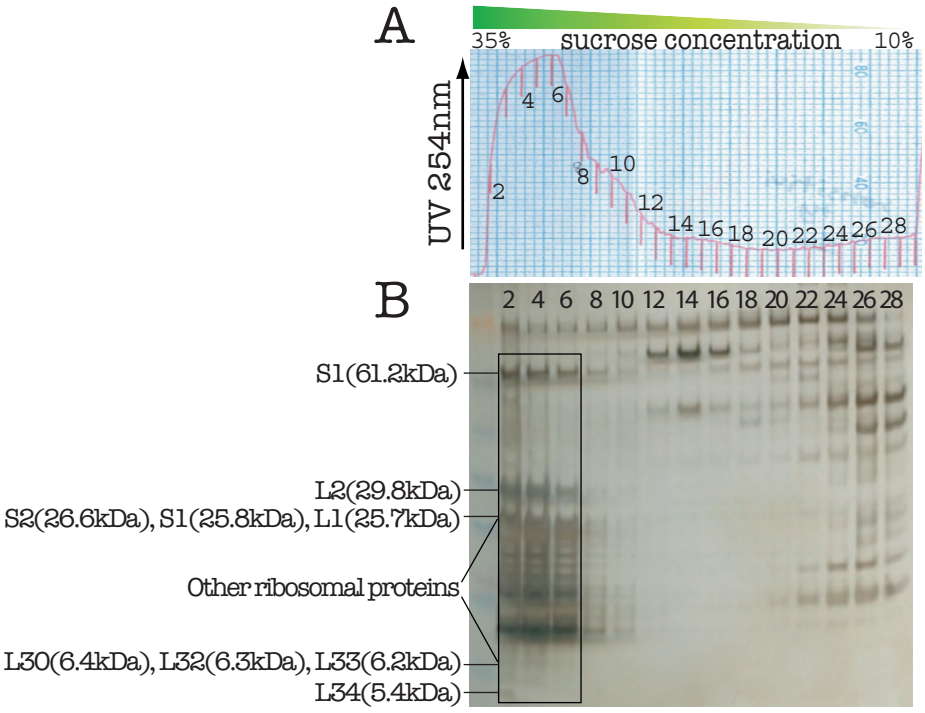

Figure 2.3: A typical 10-35 % (w/v) sucrose gradient fractionation after ultracentrifugation at 22,000 rpm for 16 hours (SW28 rotor, Beckman). The fractionation profile (A) shows the UV absorbance reading at 254 nm and the corresponding fractions are shown by SDS-PAGE in panel (B). The fractions highlighted in the box represent 70S ribosomes that are sufficiently pure for preparation of the NMR sample.

2.2.1.2 ^{1}H 1D NMR spectra of 70S ribosomes

1D ^{1}H NMR experiments of intact 70S ribosomes were recorded as described in Section 5.3.1.1 on page 203. Within the 1D profile, there are several regions that are of interest for different purposes, as highlighted in Figure 2.4. The amide/aromatic (pink box) and the

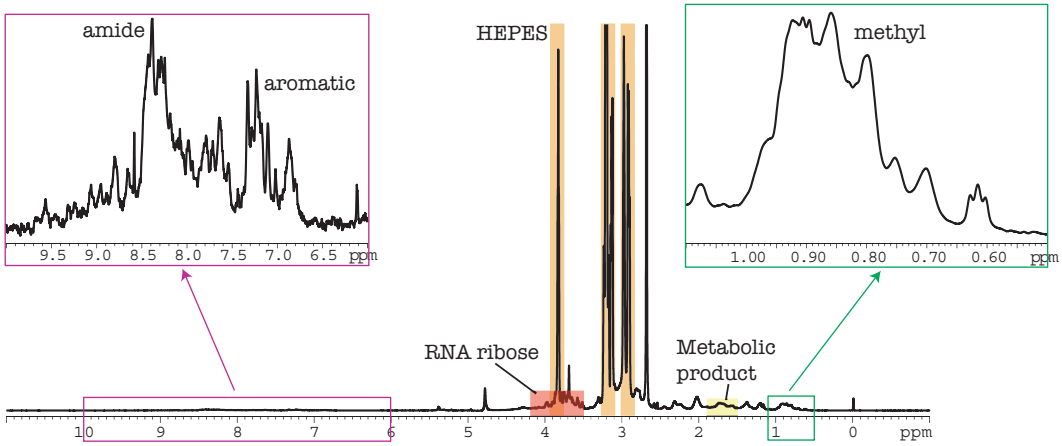

Figure 2.4: 1D ^{1}H spectrum of intact ribosomes. Amide, aromatic and methyl regions show the protein signals. The position of the HEPES peaks highlighted in orange is an indication of pH. The region highlighted in yellow is where metabolic-product signals appear as a result of microbial contamination. The RNA ribose signals highlighted in red represent the integrity of the core of the ribosome.

methyl (green box) regions are protein signals arising from the ribosome particles, and whose signal intensities are used for diffusion measurements (Section 2.2.1.5). The peaks corresponding to the HEPES present in the buffer are highlighted in orange; the change in position of the HEPES peaks is a useful indicator of any changes in pH within the sample (see Figure 2.8). Such changes in pH are linked to the presence and increase in intensity over time of metabolic products (Figure 2.4, yellow). Together these two signal regions of signals report on microbial growth contamination, which is usually associated with the degradation of the ribosome, and therefore are an important monitor of the stability of the sample over time. In addition to signals arising from proteins, ribose groups from the ribosomal RNA (rRNA) give signals that mainly fall in the region between 3.5 and 4.2 ppm. The signals from rRNA are usually very stable over time, and a change of intensity of these signals is indicative of complete degradation of the ribosome particle, however, this was found to occur very rarely.

2.2.1.3 ^{15}N-^{1}H 2D spectra of 70S ribosomes

The ^{15}N-^{1}H HSQC spectrum of uniformly ^{15}N-labelled intact 70S ribosomes is shown in Figure 2.5, displays several interesting characteristics as described previously [34]. Despite

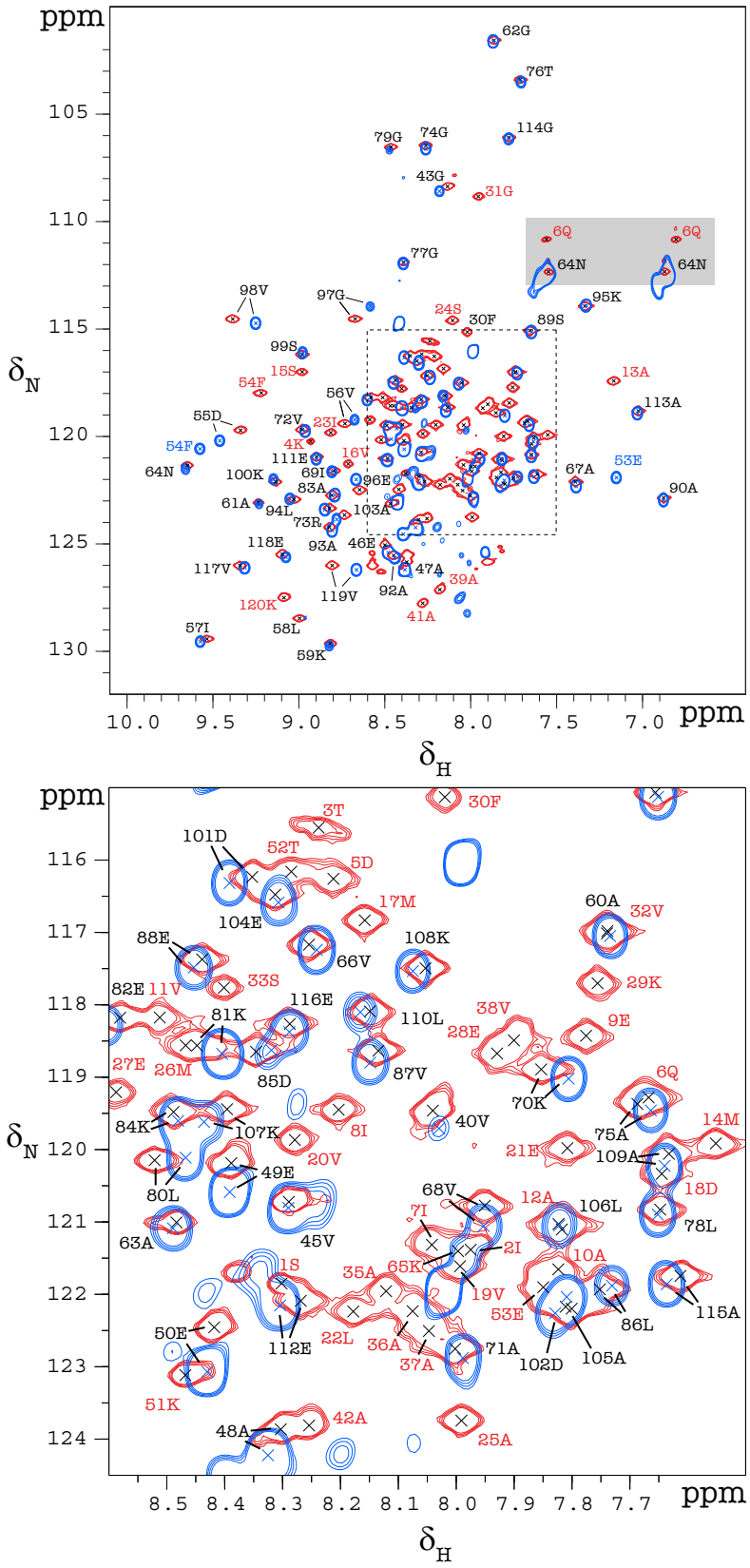

Figure 2.5: ^{15}N HSQC spectrum of 10 µM 70S ribosomes (blue) overlaid with that of isolated L7/L12 (red) recorded in Tico buffer at 25 °C. Residue assignments for peaks that closely overlay in the two spectra are in black text. Where the peak positions are different for the two species, the isolated and ribosome-bound L7/L12 assignments are in red and blue text, respectively. The bottom panel shows an expanded view of the crowded central region for clarity. Peaks in the grey box correspond to side-chain NH_2 resonances arising from the asparagine and glutamine residues.

the large size of ribosome complex (7500 amino acids), the spectrum shows a subset of ~80 well-dispersed resonances. These correspond to cross-peaks that arise from the CTD and part of the hinge region (residue 40 onwards) of the L7/L12 protein on the ribosome, which is observed in an overlay with the spectrum of the L7/L12 protein in isolation and for which its assignment is shown (Figure 2.5). These spectra demonstrate that NMR reports on a highly dynamic region of the otherwise largely inflexible ribosome complex. Apart from the cross-peaks arising from residues in L7/L12, there are also several additional peaks that are thought to correspond to protein S1 in the small subunit of the ribosome [179].

2.2.1.4 Isotopic labelling efficiency of the ribosome samples

As the ribosomes used to record the RDC experiments were uniformly ^{15}N labelled *in vivo* within *E.coli*, the efficiency of the uniform ^{15}N isotopic labelling was therefore tested by comparing the intensities of simple ^{15}N-edited and ^{15}N-filtered 1D proton spectra, as shown in Figure 2.6. The ^{15}N-edited spectrum, which only contains signals from protons attached to ^{15}N, can be acquired as the first row of a standard ^{15}N HMQC spectrum (blue). The ^{15}N-filtered spectrum, which only contains signals from protons not attached to ^{15}N (green), is then easily obtained by simple modification of the phase cycle of the ^{15}N HMQC sequence so as to reject ^{15}N-labelled magnetisation. In brief, the ^{15}N 90 ° pulses are (independently) cycled {*x,-x*} as in the standard HMQC sequence, but with no cycling of the receiver. The relative intensities of the signals in the amide region of these two spectra is then an accurate measure of the labelling efficiency. The sample used for RDC measurement was found to be more than 95 % ^{15}N-labelled as shown in Figure 2.6.

2.2.1.5 Monitoring the integrity of 70S ribosome overtime using diffusion experiments

An important tool for describing the integrity of the ribosome particles during the course of NMR data acquisition is the diffusion coefficient. As described in section 1.3.3 on page 22, the diffusion coefficient is best calculated by fitting the observed intensity profile to Equation 1.3. Ideally, measurements of intensities for a minimum of 10 different gradient

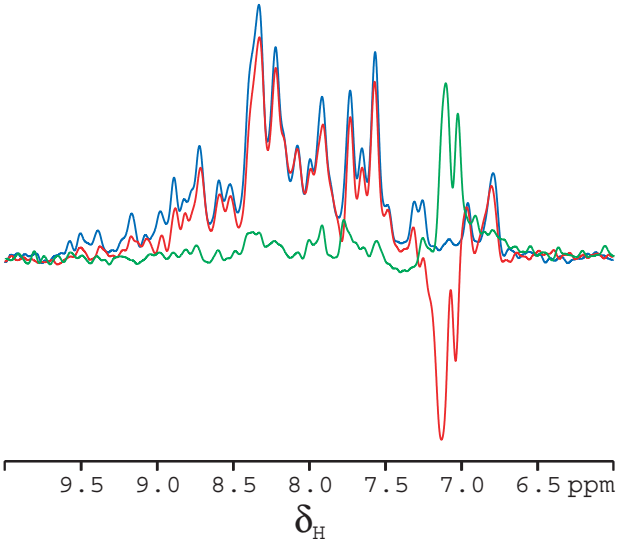

Figure 2.6: Verifying the labelling efficiency of the ribosome. The ^{15}N-edited spectrum of 70S ribosomes (blue) contains only signals from protons attached to ^{15}N. The ^{15}N-filtered spectrum (green) contains only signals from protons not attached to ^{15}N. The red spectrum contains signals from all protons, but with those from protons attached to ^{15}N appearing positive and those from protons not attached to ^{15}N appearing negative.

strengths enable good curve-fitting, but due to the limited signal and in order to maximise the acquisition time of the 2D experiments, which are collected as an interleaved manner, only three data points at 5%, 50% and 95% relative gradient strength were measured routinely. Within the multiple peaks in the spectra, only the amide and methyl regions were used to calculate the diffusion coefficient (as shown in Figure 2.7 A). The total signal intensity for each gradient strength is calculated by integrating the spectra over the selected region. Instead of fitting the three intensities to a curve described by Equation 1.3, three values of the diffusion coefficient are calculated from the pairwise intensity ratios for the different gradient strengths (*i.e.* $I_{95\,\%}/I_{50\,\%}$, $I_{50\,\%}/I_{5\,\%}$, $I_{95\,\%}/I_{5\,\%}$) according to:

$$D_{I_B/I_A} = \frac{\ln(I_A/I_B)}{(\Delta - \delta/3) \cdot (\gamma\delta)^2 \cdot (G_B^2 - G_A^2)} \tag{2.1}$$

where $G_B > G_A$.

The three diffusion coefficients calculated in this manner give an indication of the heterogeneity of the diffusion profile, as well as a single average diffusion coefficient. For an intact 70S ribosome sample, the measured diffusion coefficient is $\sim 2\times10^{-11}$ m^{2}s^{-1}, changes

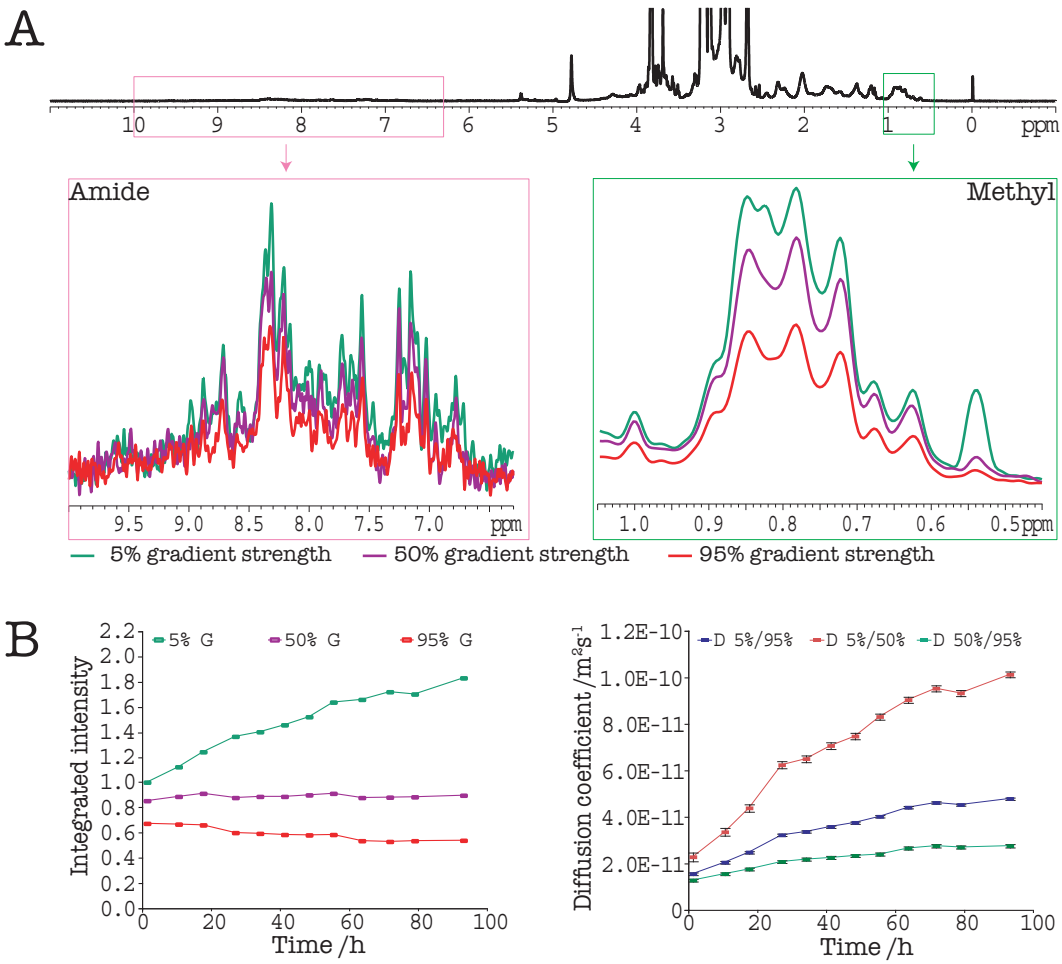

Figure 2.7: **A:** Amide (left) and methyl (right) regions of the ^{1}H ribosome spectrum (above)showing diffusion-mediated intensity decay for three gradient strengths. The signal-to-noise ratio is best for the methyl region. **B:** Plot of integrated intensity of each data point (left) and the calculated diffusion coefficients from the pairwise intensity ratios of the different gradient strengths.

observed within diffusion coefficient is an indication of degradation of the ribosome.

Together with measuring the diffusion coefficient, the diffusion properties of the sample are evaluated over time, to gauge the integrity of the sample. Figure 2.7 B shows an example demonstrating typical changes in diffusion over time for a ^{13}C,^{15}N uniformly labelled ribosome sample. At time zero, the three values for the diffusion coefficient calculated from the three pairs of gradient strengths are very similar, reflecting the high degree of sample homogeneity. Over a time period of 100 hours, the measured diffusion coefficient begins to increase (Figure 2.7 B), with the value calculated from $I_{50\%}/I_{5\%}$ increasing more rapidly than that from $I_{95\%}/I_{50\%}$, which remains close to $2{\times}10^{-11}\,\mathrm{m^2 s^{-1}}$. This discrepancy

between the calculated values is a reflection of increasing heterogeneity within the sample. The diffusion coefficient calculated from the $I_{50\%}/I_{5\%}$ ratio is biased towards small species that have become detached from the ribosome, while that calculated from the $I_{95\%}/I_{50\%}$ is closer to the diffusion coefficient of the ribosomal core particle.

The increasing heterogeneity is also apparent from the plot of the integrated intensities for the three gradient strengths (Figure 2.7 B). These plots show that the increase in the diffusion coefficient calculated from the $I_{50\%}/I_{5\%}$ is dominated by increasing intensity in the 5 % spectrum. This suggests a contribution to the integrated signal region from small species that were not NMR-visible at time zero. Overall, the evolution of the diffusion profiles is consistent with a picture of sample-ageing in which the majority of the ribosome particle remains intact, but with the progressive detachment of some small peripheral proteins. However, it must be noted that this occurs over a very slow time period of ~100 hours, which therefore permits the extended acquisition of data.

2.2.1.6 Monitoring the integrity of 70S ribosomes over time using ^{1}H 1D analysis

In addition to the use of diffusion experiments, the integrity of the 70S ribosomes is also monitored using 1D spectra. As discussed in section 2.2.1.2, four characteristic regions are followed for this purpose (Figure 2.8, top), corresponding to HEPES, metabolic products, methyl groups and RNA ribose signals. The change of the HEPES peaks is an indication of changing pH, which becomes lower if the sample is degrading [207]. This is also found to coincide with the appearance of a signal at ~1.8 ppm, which is suspected to be a metabolic byproduct resulting from microbial growth [207]. For example, in the ^{15}N 70S ribosome sample used for the measurement of N-H J-couplings, the change of the HEPES chemical shift is very small (<0.1 ppm, Figure 2.8 A), even though increased intensity was observed at the chemical shift suspected to correspond to a metabolic product (Figure 2.8 B). The signals arising in the methyl region (0.9 ppm) are used as a representation of ribosomal protein integrity (Figure 2.8 top, green) whereas the signals arising from RNA ribose group are a representation of the integrity of the RNA molecules within the ribosome (Figure 2.8 top,

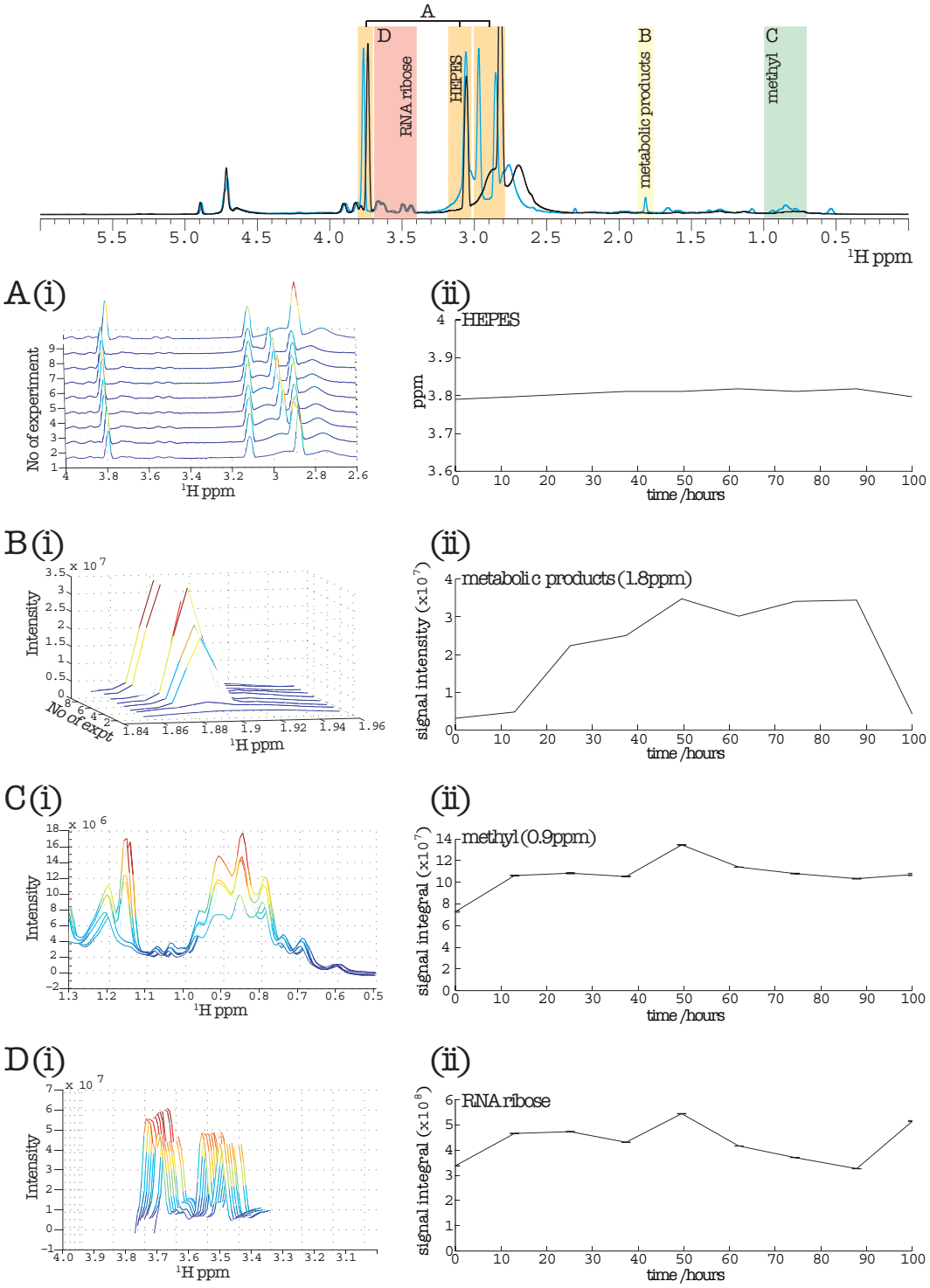

Figure 2.8: Top: ^{1}H 1D spectrum of a ^{15}N 70S ribosome sample recorded at time=0 (black) and at time=100 hours (blue). The HEPES, RNA ribose, metabolic product and methyl regions are highlighted in orange, red, yellow and green, respectively. **A, B, C** and **D:** Stack of ^{1}H spectra of ^{15}N ribosomes recorded over increasing sample lifetime at 25 °C, zoomed to show the regions corresponding to (A-i) the HEPES, (B-i) the metabolic byproduct, (C-i) the methyl groups and (D-i) the RNA ribose signals. The chemical shift change of the downfield HEPES peak is shown in A(ii), and the evolution of the signal intensities in the metabolic byproduct, methyl and RNA-ribose regions are shown in B(ii), C(ii) and D(ii), respectively.

red). In this example, both the methyl and RNA signals are shown to be stable over time with minor fluctuations arising due to spectra noise (Figure 2.8 C&D). This suggests that the 70S ribosome remains intact.

2.2.1.7 ^{15}N spin relaxation of the 70S ribosome

To extend the previous studies which examined the dynamical properties of 70S ribosome [34], backbone ^{15}N relaxation rates were measured for the isolated L7/L12 protein and for the 70S ribosomes (recorded by Dr Helene Launay, UCL).

The transverse relaxation rate (R_2) is the most easily interpretable measure of the dynamic properties of the protein. The R_2 rates for the isolated L7/L12 protein revealed that the NTD was the fastest-relaxing region of the protein (average 17 Hz), with the CTD signals relaxing slightly slower (average 13 Hz). The hinge region has the slowest relaxation rate (average 3 Hz), as it is largely unstructured and therefore has the greatest local flexibility (Figure 2.9). The backbone ^{15}N relaxation properties of isolated L7/L12 are discussed in more detail in Section 2.2.3.4.

The R_2 of L7/L12 bound to the ribosome was only measured for the CTD residues as the NTD is the interaction site with the ribosome and is not visible in 2D HSQC spectra. The relaxation data recorded on the ribosome was noisier than that for the L7/L12 protein as the signal-to-noise ratio is limited by the low concentration of the ribosome sample. Generally, it was found that the average R_2 value (17 Hz, indicated by red line) is larger than for the isolated L7/L12 protein in solution, suggesting that the L7/L12 CTD is less mobile on the ribosome than it is in solution, despite the flexibility of the long linker that tethers the CTD to the ribosome via the NTD. This suggests either that the ribosome has an intrinsic influence on the dynamic properties of L7/L12, perhaps through specific interactions, or that the linker is not sufficiently flexible to render the motion of the CTD independent of the NTD and ribosome.

The longitudinal relaxation rate (R_1), shows a decrease for the L7/L12 bound to the ribosome compare to the isolated L7/L12 protein. This is in agreement with the increase

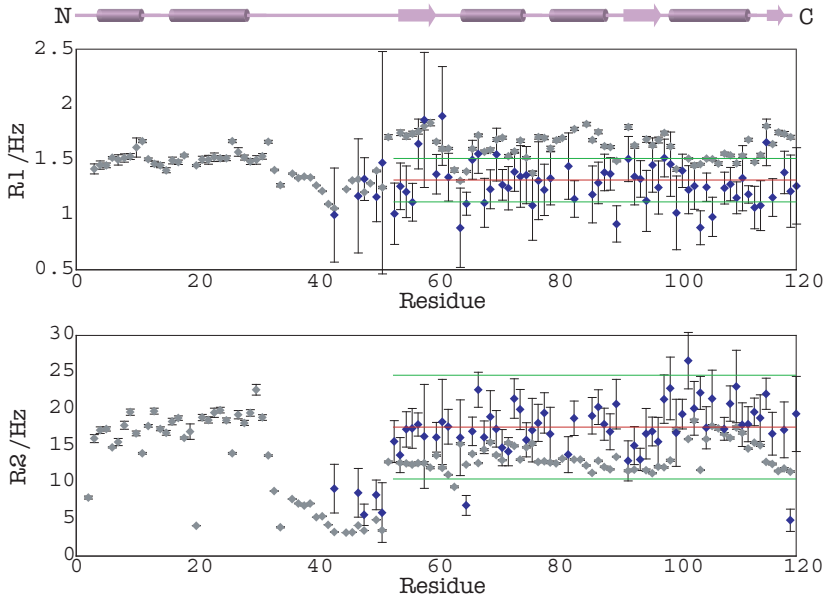

Figure 2.9: Plot of R_1 and R_2 relaxation rates of the isolated L7/L12 dimer (grey diamonds) and that bound to the ribosome (blue diamonds). The means of the relaxation rates of the CTD residues of L7/L12 bound to the ribosome are indicated with red lines, and the means +/- uncertainties are indicated with green lines. The secondary structure of L7/L12 is shown above.

in the transverse relaxation rate (R_2), since the corresponding rotational correlation time is in the R_1 regime where the R_1 relaxation rates decrease with increases in the rotational correlation time.

2.2.2 Characterisation of isolated L7/L12

2.2.2.1 Expression and purification of recombinant L7/L12 protein

To compare the structural characteristics of L7/L12 both in the ribosome and in solution, the L7/L12 protein was expressed and purified as described in sections 5.1.10 and 5.1.11. The protein localised to the soluble fraction and expressed in significant quantities, typically 12 mg of material was recovered from a one litre culture. It was found that during the final purification step using gel filtration chromatography (Superdex 200), the elution profile showed multiple peaks (Figure 2.10 A), eluting at 60 ml (peak 1), 80 ml (peak 2), 115 ml (peak 3) and 127 ml (peak 4). SDS-PAGE of the fractions suggested that all peaks contain the same protein with molecular weight corresponding to L7/L12, which in turn suggests

that all peaks in the gel-filtration trace represent molecular species composed of the same protein, *i.e.* different conformational and/or oligomeric states. Peaks 1 and 2 were subjected to further investigation to identify the different species, while Peaks 3 and 4 contained too little material for further analysis.

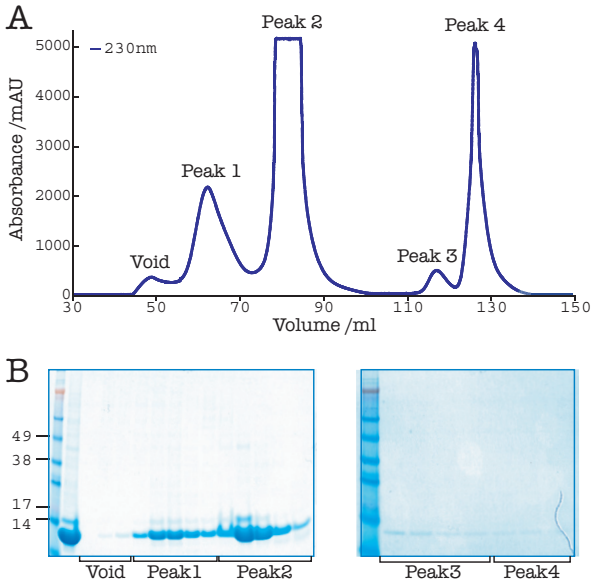

Figure 2.10: A: Gel filtration elution profile of L7/L12 and **B:** SDS-PAGE analysis of fractions from corresponding peaks. The elution profile of the protein was monitored using the backbone UV absorbance at 230nm as the L7/L12 protein does not contain any aromatic residues.

2.2.2.2 Investigation of the different species observed by gel filtration chromatography

As discussed in the previous section, multiple peaks were observed on the gel filtration profile during the purification of the isolated L7/L12 proteins. The observation of only L7/L12 protein by SDS-PAGE indicated that there is no co-purification of L10, nor other impurities. The most possible explanation for the multiple peaks present in the gel filtration profile represent different oligomeric states of the protein. Following purification, peak 1 and peak 2 (Figure 2.10) were pooled and concentrated separately as two fractions, fraction 1 and fraction 2, respectively.

Initial ^{15}N diffusion analysis To further investigate the behaviour of fraction 1 and 2, the diffusion coefficients were measured for both fractions. The measured diffusion coefficient of fraction 1 ($5\times10^{-11}\,m^2s^{-1}$) is smaller than fraction 2 ($7.3\times10^{-11}\,m^2s^{-1}$), which is in

agreement with the observation of the gel filtration profile. If the species corresponding to these two fractions are slow-exchanging oligomeric forms, the expectation is a resulting inhomogeneous diffusion intensity profile, and a diffusion coefficient calculated from a single Gaussian fit that is inversely related to the protein concentration. However, the diffusion intensity profiles were found to be homogeneous and well-fitted by a single Gaussian. Moreover, the calculated diffusion coefficient for fraction 1 was smaller than fraction 2, despite being lower in concentration(160 μM vs 240 μM). Together, the diffusion data indicates no exchange between the species of fraction 1 and fraction 2.

Analytical gel filtration analysis The two fractions were then applied to an analytical gel filtration column (Superdex 10/300 GL) to reassess the conformational and oligomeric states of the protein. In addition, a sample of the L7/L12 protein purified from a ion-exchange column but without further purification by gel filtration was also applied to serve as a control for the appearance of fractions 1 and 2.

As shown in Figure 2.10 A, the elution profile of the control sample reproduces the original profile of the preparative column, showing two major protein-containing peaks, one eluting at 12 ml (peak 1, red star) and another at 15 ml (peak 2, green star).

When fraction 1 was subsequently applied to the column, the elution profile showed two peaks, peak 1, at an early elution volume of 12 ml (red star, Figure 2.11 B) and another at a slightly later elution volume of 16.5 ml (green star, Figure 2.11 B). It was noted that peak 2 eluted slightly later than in the control sample. The relative proportions of the two peaks therefore, as calculated by integrating the area beneath the curves, suggested that fraction 1 is mainly composed as peak 1(> 85 %) and a small fraction converted to peak 2 (< 15 %). Therefore compared to the control sample, the molecular species present in fraction 1, is relatively stable and does not convert into peak 2 easily. In contrast to fraction 1, when fraction 2 was applied to the column, it gave rise to a single species eluting at *ca.* 15 ml, which corresponds to peak 2. These data confirm that the conformational state adopted by fraction 2 is an inherently stable molecular species.

To evaluate the conformational state of the L7/L12 further, the control sample (as

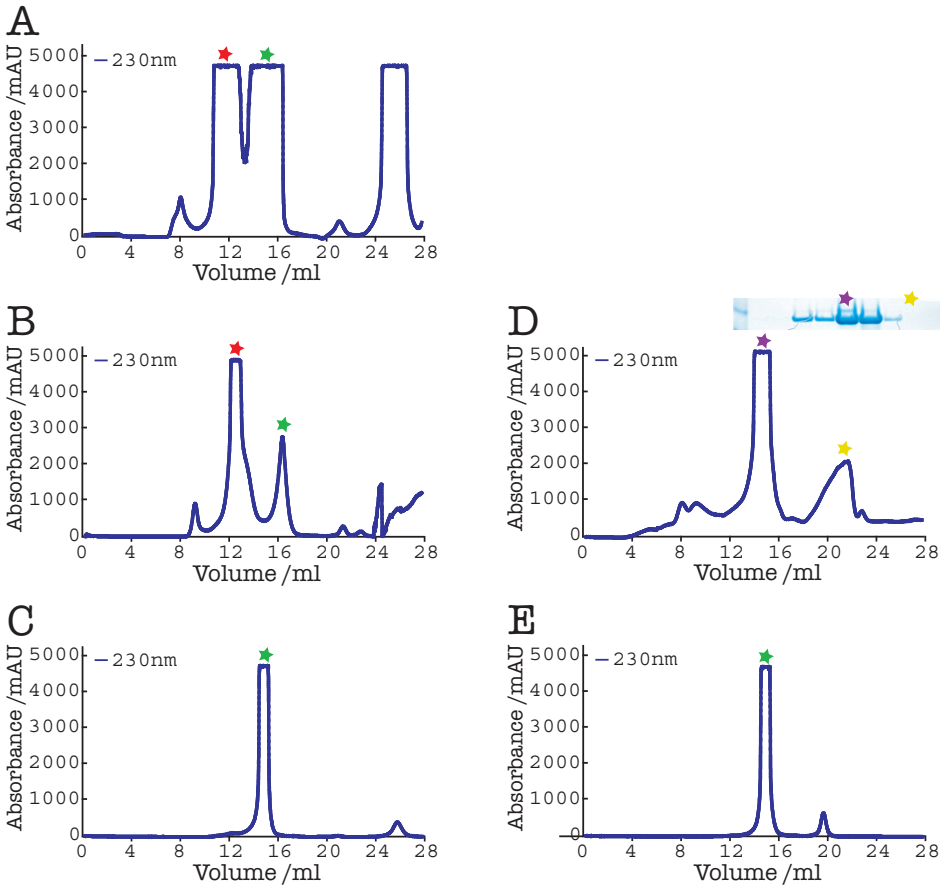

Figure 2.11: Analytical gel-filtration of the L7/L12 protein on Superdex 10/300 GL column. **A:** Sample prior to gel filtration loaded as reference. **B:** Peak 1 (red star) from A loaded, resulting in two peaks: ~85% peak 1 and ~15% peak 2. **C:** Peak 2 (green star) loaded, resulting in the same peak. **D:** Total sample denatured and loaded onto column pre-equilibrated in 8M urea, resulting in one peak. The peak highlighted with the yellow star showed no band on gel. **E:** Profile generated by loading a sample prepared by refolding the protein-containing peak in D (purple star) in native buffer, showing a single peak at the same position as peak 2.

described previously above) was unfolded in 8M urea and applied to the column under the equivalent denaturing conditions. It was found that the protein eluted at a single peak of volume of *ca.* 15 ml (Figure 2.11 D, purple star). This suggested that L7/L12 assumed a single, presumably unfolded conformation. When the 8M-denatured control sample was refolded and applied to the column under native conditions, a single peak eluting at 15 ml was observed (Figure 2.11 C, green star). Both the denatured and refolded samples were examined by ^{1}H 1D NMR spectroscopy to ensure the completeness of denaturation and refolding (see Appendix A, Figure A.1). The spectrum recorded on the refolded sample is very similar to folded L7/L12 spectrum whereas the spectrum recorded on the denatured

sample shows a typical unfolded spectrum. In addition, the diffusion intensity profiles for both samples indicates homogeneous and the diffusion coefficient measured on the refolded sample is same as that measured on fraction 2, with same concentration.

Together, the data suggest that the L7/L12 can adopt different isoforms that have non-globular characteristics, as can be seen by the similar elution volumes of the 8M denatured L7/L12 (Figure 2.11 D) and the elution volumes of peaks 1 and 2 in the native from; this is due to the architecture of the protein in which there is significant degree of disorder (ca. 20 % of the sequence) that arises from the hinge regions between the NTD and CTD. Also, the refolding data indicate that under certain conditions, L7/L12 can also acquire a uniform conformation, which appears to have characteristics that resemble fraction 2. Moreover, this refolding study also indicated that only the species in fraction 2 can be generated *in vitro*, but not fraction 1. The non-globular characteristics also prevent a more detailed quantitative estimate of the relative sizes of the isoforms present in fractions 1 and 2. As discussed in Section 2.1.2, studies have suggested that dominant conformation of L7/L12 is that of a dimer, but it can also form higher order oligomers such as tetramers. Based upon these previous studies and the observations in this study, it would suggest that fraction 1 of L7/L12 (the earlier eluting species) most likely represents a tetrameric species and fraction 2 represents the dimer, with no exchange between these two species. Bearing these characteristics in mind, 2D NMR spectra were recorded for both species.

For the following characterisation of L7/L12 by NMR and subsequent RDC experiments presented later in this Chapter and beyond, unless otherwise indicated, these two species are referred as fraction 1 (early-eluting fraction, higher order oligomer) and fraction 2 (later-eluting fraction, lower order oligomer).

^{15}N-^{1}H 2D NMR characterisation To further investigate the difference between fractions 1 and 2, 2D ^{15}N HSQC spectra were recorded on both fractions (Figure 2.12). The peaks from residues in the CTD overlay between the two samples and appear with similar intensities (after scaling for differences in concentration). However, the recognised peaks of the NTD appear with much weaker intensity in the sample of fraction 1 (Figure 2.12, red

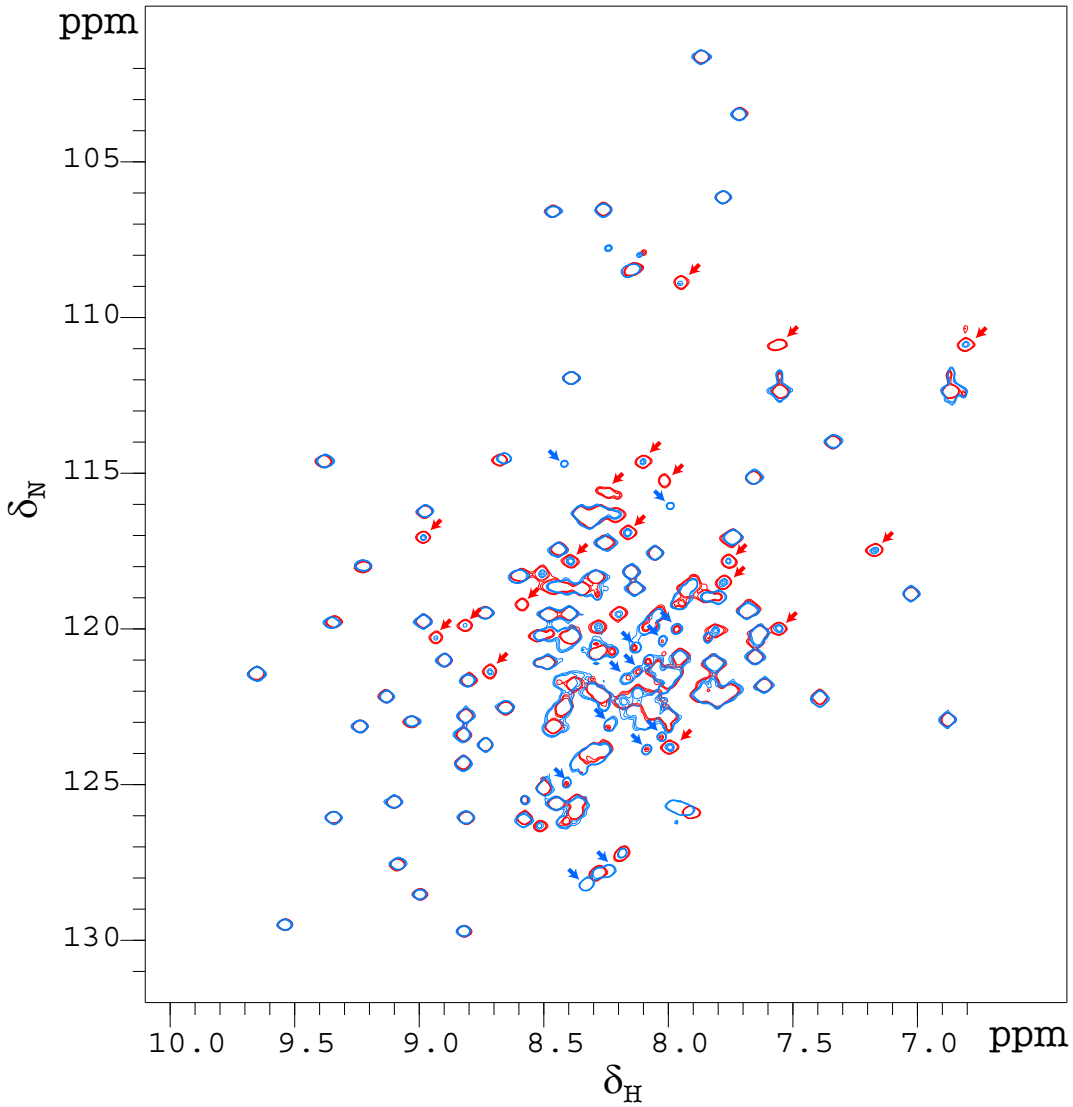

Figure 2.12: Overlay of ^{15}N HSQC spectra of fraction 1 (blue) and fraction 2 (red) of isolated L7/L12. Red arrows indicate recognised NTD peaks, which are weak in the fraction 1 spectrum. Blue arrows indicate unassigned peaks that are present with stronger intensity in the spectrum of fraction 1.

arrows) . This spectrum also contains ~25 peaks that do not correspond to any of the known assignments (Figure 2.12, blue arrows). This subset of unassigned peaks also appear in the spectrum of fraction 2, but with a much weaker intensity.

One explanation for these observations is that the species in fraction 1 has a different structure at its NTD, which is reflected by a different pattern of NTD peaks in the ^{15}N HSQC spectrum. The weak peaks appearing at the recognised NTD shifts could be due to a small population of the species corresponding to fraction 2. Similarly, the weak peaks appearing at the unrecognised positions in the spectrum of fraction 2 may be the result of low-level contamination of this sample by the principal species of the other fraction. In summary, the NMR spectra show that the species in the two fractions appear to differ only at their NTDs. Together with the observations from the initial diffusion experiments and analytical gel filtration, the conclusion is that fraction 1 and fraction 2 are non-interconverting forms of L7/L12 that differ in their conformation and/or oligomerisation state at their NTDs.

Further evidence for an oligomerised state was observed during the detailed inspection of 2D ^{15}N HSQC spectra of fraction 2, it was observed that the linewidths of the recognised NTD peaks were noticeably broader than those of the CTD peaks, despite the smaller molecular weight of the dimeric NTD. This suggested that further oligomerisition of the dimeric NTD may be occurring, as suggested previously [202–204].

Evalution of the oligomeric states of the L7/L12 protein in fraction 2 using ^{15}N diffusion experiments To investigate the possible higher order oligomerisation within fraction 2, the diffusion coefficient of fraction 2 was measured as a function of L7/L12 monomer concentration using ^{1}H diffusion NMR experiments (Section 5.3.1.4). From the measured diffusion coefficients (D), the hydrodynamic radii and molecular weights are estimated as described in Section 1.3.3.

As shown in Table 2.1, the estimated molecular weights at all four concentrations correspond to values that are significantly larger than both the molecular weight of a monomer (~13.4 kDa) or that of a dimer (~27 kDa). This is at least partly due to the structure of the L7/L12 protein (Figure 2.2 B&D), which is two folded domains connected by a flexible

linker (*i.e.* significantly different from a spherical shape. As the overall shape of the protein is not expected to be compact, the approximation of the protein as a sphere (as assumed for the MW calculations) is inaccurate, which means that the calculated molecular weights represents upper bounds for the true molecular weight.

Concentration /μM	D /m^2s^{-1}	r_h /nm	Estimated M_r /kDa
650	6.53×10^{-11}	3.74	153
240	7.3×10^{-11}	3.35	108
70	7.73×10^{-11}	3.16	90
18	8.86×10^{-11}	2.76	58

Table 2.1: Measured diffusion coefficient of L7/L12 as a function of monomer concentration and corresponding hydrodynamic radii and estimated molecular weights.

To obtain a realistic estimation of the expected diffusion coefficient for L7/L12 from the diffusion coefficient, the hydrodynamic radius and diffusion coefficient of the protein were calculated theoretically from the available NMR structure. Using this strategy, two species were considered, the monomeric and dimeric forms of the protein. As shown in Figure 2.13, three PDB files of these two species were derived from the NMR solution structure of L7/L12 (source PDB 1RQU [22]), which represents the dimeric state. These structures were then used to calculate the hydrodynamic properties of the different species using HYDROPRO [208], including the hydrodynamic radii and diffusion coefficients.

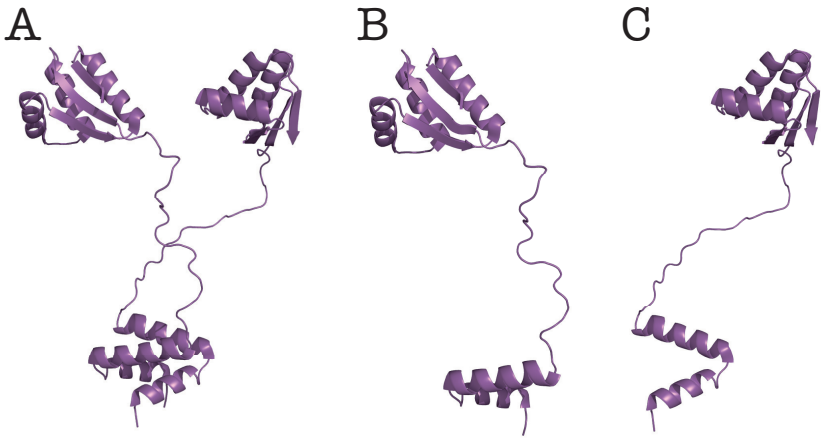

Figure 2.13: The different states of the L7/L12 protein used for HYDROPRO calculations. **A:** Dimeric form with both hinge region extended. **B** and **C:** each of the two monomers derived from the dimer PDB. PDB ID: 1RQU [22]

The results from HYDROPRO are summarised in Table 2.2, and these were then

used to reconsider the experimentally measured diffusion coefficients in Table 2.1. The measured diffusion coefficient of L7/L12 at the highest monomer concentration (650 μM), was 6.53×10^{-11} m^2s^{-1}, corresponding to slower diffusion than that predicted for the dimer by HYDROPRO (7.55×10^{-11} m^2s^{-1}). This suggests the presence of higher order oligomers in the 650 μM sample. Conversely, at the lowest monomer concentration (18 μM), the measured diffusion coefficient (8.86×10^{-11} m^2s^{-1}) is faster than that predicted for the dimer, suggesting that the monomer is populated to a significant degree at this concentration. At a monomer concentration of 70 μM, however, the measured diffusion coefficient is similar to that predicted for the dimer.

Input	r_h / nm	$D/10^{-11}$ m^2s^{-1}
dimer	3.236	7.552
monomer [A]	2.642	9.252
monomer [B]	2.643	9.248

Table 2.2: Summary of results from HYDROPRO, showing estimated hydrodynamic radii and calculated diffusion coefficients of different L7/L12 conformations.

Overall, the measured diffusion coefficients in comparison to those calculated theoretically suggest that monomeric, dimeric and tetrameric forms of L7/L12 are present over the measured concentration range (18-650 μM). However, this interpretation should be treated with caution, as the diffusion coefficient of the monomeric and dimeric states calculated using HYDROPRO are only estimated. This is because the state of L7/L12 in solution will be an ensemble of multiple conformations in the linker region, rather than being fixed in a single conformation as used for the HYDROPRO calculations. Nevertheless, the observed increase in measured diffusion coefficient with decreasing L7/L12 concentration unambiguously identifies a change in the oligomeric state over this concentration range. In addition, the observation that all the diffusion-mediated intensity profiles were well fit by a single Gaussian indicates that the different oligomeric states are in fast exchange on the diffusion timescale (100 ms).

In principle, it should be feasible to use the concentration-dependent diffusion coefficient to determine the dissociation constant of the monomer-dimer and proposed dimer-tetramer

exchange processes. However, this would be difficult if the individual diffusion coefficients of the of the different oligomeric states are not known. Determining these diffusion coefficients is likely to be challenging if the dissociation constants for the two processes are not very different.

2.2.2.3 Summary

Combining all the results from the analysis above, it can be concluded that only the L7/L12 protein is present in the two fractions obtained by gel filtration, and that the species contained within these two fractions are non-exchanging. The species represented by fraction 2 can be reformed by unfolding/refolding *in vitro*, but not the species represented by fraction 1. Therefore the initial hypothesis that fraction 1 is a tetrameric state in slow exchange with the dimeric species of fraction 2 must be discounted. Indeed, it was shown by concentration dependent diffusion measurements that the different oligomeric states (monomer, dimer, tetramer) are present within fraction 2, and are in fast exchange with each other. The nature of the species in fraction 1 remains unknown. It diffuses more slowly than the species in fraction 2, even at much lower concentrations and gives a ^{15}N HSQC spectrum that shows differences to that of fraction 2 for the NTD residues. In addition, it appears to be stable, with only limited conversion into fraction 2, and is not formed from fraction 2 *in vitro*. Overall, these observations are suggestive of a locked higher-order oligomeric state, bound together by interactions between NTDs that adopt a different conformation to that found in fraction 2.

2.2.3 Analysis of N-H RDCs measured in phage

2.2.3.1 Alignment using bacteriophage

To be able to measure RDCs, partial alignment of the sample is required. In this case, the partial alignment of the ribosome particle was achieved using Pf1 phage (ASLA) as the alignment medium. For initial evaluation, a sample was prepared of ^{15}N-labelled 70S ribosomes with phage at similar concentrations to those required for NMR experiments (10 µM 70S ribosomes with 10 mg ml^{-1} phage). The sample was observed under a light microscope and no visible precipitation or aggregation was present under these conditions, suggesting that the phage had no immediate negative impact on the ribosome.

For the measurement of RDCs, an appropriate concentration of phage is required to provide a suitable degree of partial alignment. Although the concentration of the phage particles can be measured by their absorbance at 270 nm, the high viscosity of the phage means that the pipetting of the phage solution is not very accurate. Instead, the concentration of phage was estimated indirectly by measuring the splitting of the deuterium signals arises from the D_2O in the buffer. There is a linear relationship between the phage concentration and the size of the deuterium splitting above 10 mg ml^{-1} phage, allowing the concentration of the phage stock to be determined and then added to the ribosome sample accordingly.

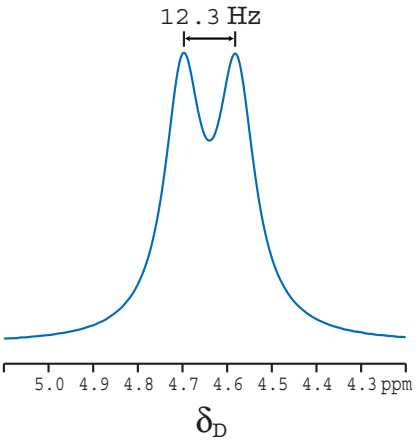

Figure 2.14: 1D ^{2}H spectrum of 70S ribosome aligned in Pf1 phage on the 700MHz spectrometer, showing splitting of 12.3 Hz at 298 K.

For this particular sample, 12.3 Hz deuterium splitting (Figure 2.14) was detected as a result of addition of phage to a final concentration 10 mg ml^{-1}.

2.2.3.2 Strategy of RDC data measurement for the 70S ribosome

Backbone amide N-H splittings were measured for the two fractions (fraction 1 and fraction 2) of isolated L7/L12 and ribosome-bound L7/L12. For each species, data were acquired on

isotropic and aligned samples (in phage) under identical experimental conditions.

Splittings for the isolated protein were recorded using an IPAP-^{15}N-HSQC experiment [157], as described in section 1.4.4 on page 34. Briefly, the N-H coupling is allowed to evolve during the ^{15}N indirect chemical shift evolution period, leading to doublets for each backbone amide in the resulting spectrum. To prevent the increased spectral crowding that would result from doubling the number of peaks, two sub-spectra are recorded in which the splitting is either in-phase or anti-phase. Subsequent addition and subtraction of the in-phase and anti-phase sub-spectra yields two further sub-spectra containing only the upfield or downfield doublet component. The splittings are extracted simply by measuring the chemical shift differences between the upfield and downfield components for each amide.

In contrast, splittings for ribosome-bound L7/L12 were measured using the HSQC/TROSY approach, where the splitting for an amide site is extracted as twice the ^{15}N frequency difference between the ^{15}N HSQC and ^{15}N TROSY-HSQC peaks. In principle, the relative error in the resulting splitting is twice that measured from the IPAP-^{15}N HSQC spectrum, assuming that the peak positions are defined with the same precision in both experiments. However, in the IPAP experiment, the upfield doublet component has a broader ^{15}N linewidth (and hence also weaker intensity) than the downfield component due to the TROSY effect (Figure 2.15). Therefore, the ^{15}N frequency of the upfield component cannot be measured to the same precision as that of the downfield component (*e.g.* residues 113A, 53E, Figure 2.15). Indeed, for some peaks with lower intensities, the upfield component broadens beyond detection (less intense than the noise), resulting in loss of information for that residue. The differential line-broadening increases as the tumbling rate of the protein slows, and it becomes beneficial to switch to the HSQC/TROSY approach for measurement of the splittings. The ^{15}N linewidth of the HSQC-peak will also be broader than that of the corresponding TROSY-peak (Figure 2.15), but the effect is not as pronounced as for the upfield and downfield components of the ^{15}N doublet, and the improvement for the HSQC/TROSY approach outweighs the disadvantage of halving the measured frequency difference [209], as shown in Figure 2.15.

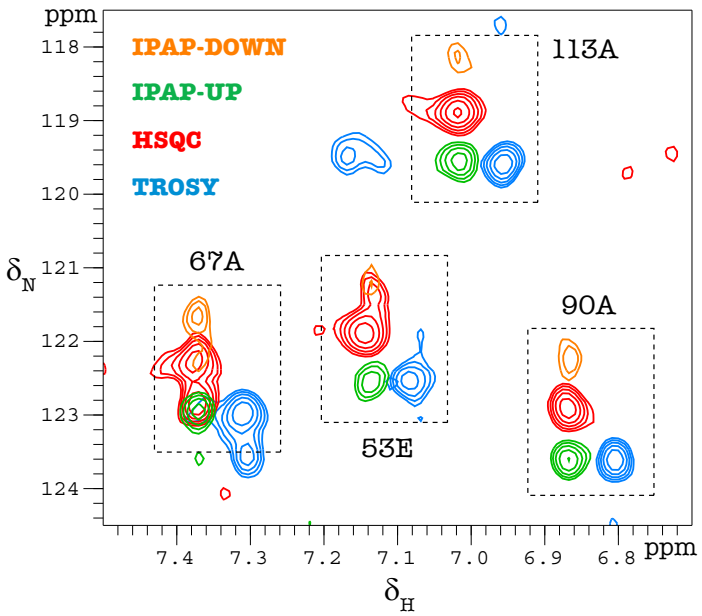

Figure 2.15: Comparison between IPAP-^{15}N-HSQC spectra (orange and green) and ^{15}N-HSQC and ^{15}N-TROSY spectra (red and blue) acquired on 70S ribosomes under the same conditions. The spectra were recorded for the same total experiment time and contoured at the same level (4× RMS noise). A selected region is displayed, containing cross-peaks from residues 53E, 67A, 90A and 113A.

Part of the motivation behind measuring RDCs for L7/L12 was to develop and test the techniques required for RNCs (chapter 3), and the comparison between the IPAP-HSQC and HSQC/TROSY approaches for the ribosome-bound L7/L12 indicated that the HSQC/TROSY approach would also be more appropriate for RNCs. Furthermore, for RNCs, the difference in the relaxation properties between the isolated and ribosome-bound protein is greater than that for isolated and ribosome-bound L7/L12, with the effective tumbling rate of the nascent chain typically much slower than that of the corresponding isolated protein. Hence the HSQC/TROSY method will be clearly required for RNCs.

The sample and experimental conditions for each of the samples are listed in table 5.9 on page 200. For the ribosome-bound samples, data were acquired by recording a repeated series of relatively short experiments in an interleaved manner. This approach allows the use of 1D and diffusion experiments to monitor the ribosome integrity at regular intervals. The final spectra for splitting measurements were obtained by summing the individual experiments over the time period for which the ribosome is deemed to be intact, as judged by diffusion experiments (diffusion data shown in Appendix A, Figure A.2).

2.2.3.3 Initial analysis of the RDCs measured on isolated and ribosome-bound L7/L12

As mentioned, the N-H RDCs of the isolated L7/L12 protein were recorded using the IPAP approach [157]. The upfield and downfield sub-spectra obtained for L7/L12 (Figure 2.16 A) show that most residues give cross-peaks in both the isotropic and aligned sub-spectra (the measured RDCs versus residue number are shown in Appendix A, Figure A.3). The proton linewidths of the peaks in the aligned sub-spectra are typically broader than in the isotropic sub-spectra, which is due to a contribution from unresolved proton-proton splittings.

Four spectra were recorded for measuring the N-H RDCs of L7/L12 on the ribosome using the HSQC/TROSY approach, which are the isotropic HSQC and TROSY spectra and aligned HSQC and TROSY spectra. As shown previously in Figure 2.5, 75 peaks were observed from ribosome-bound L7/L12 (NTD peaks absent) in the isotropic spectra. Of the 75 peaks observed in the isotropic spectra, 70 peaks were observed in the aligned spectra. The reduced number of peaks is due to the overall broadening of signals in the aligned spectra, which causes some weak peaks to broaden beyond detection. From the isotropic HSQC and TROSY spectra, the J couplings were calculated by taking the differences of the ^{15}N frequency for pairs of HSQC and TROSY peaks in Hz. The same set of differences calculated from the aligned HSQC and TROSY spectra, then yields the J+D splittings. The D value was then calculated by taking the difference between the J and J+D values (Figure 2.16 B, top right panel). All four sets of spectra that were used to measure the RDCs are shown in Figure 2.16 B, with an example of the dipolar couplings obtained for a set of cross-peaks shown (right, bottom panel). 70 RDCs were measured, and of these data, resonances that were overlapped were discarded, resulting in 60 RDCs, which were derived almost exclusively from the CTD (the measured RDCs versus residue number are shown in Appendix A, Figure A.3).

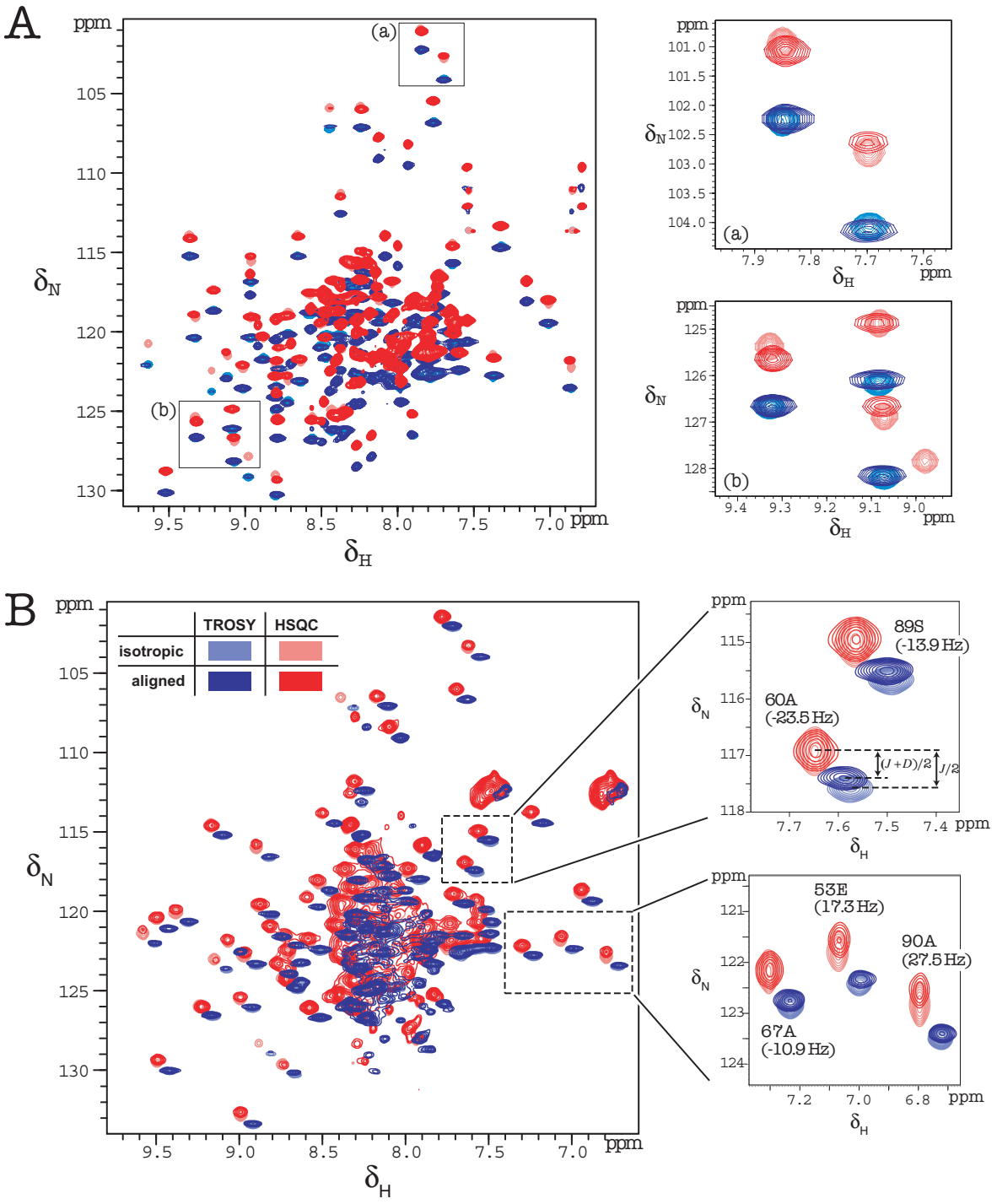

Figure 2.16: A: Overlay of isotropic (pale red and blue) and anisotropic (dark red and blue) ^{1}H-^{15}N IPAP spectra of uniformly ^{15}N-labelled isolated L7/L12 protein. **B:** Overlay of isotropic (pale red and blue) and anisotropic (dark red and blue) ^{15}N HSQC and TROSY spectra of uniformly ^{15}N-labelled *E. coli* 70S ribosomes.

2.2.3.4 Considerations for data selection used in structural analysis

In order to interpret RDC values and apply them to extract structural information, a more detailed analysis of the dynamic properties of L7/L12 protein is needed. While the scope of the current work was restricted to analysing the RDCs from L7/L12 with respect to the structure and orientational probability distribution of the domains, with application of RDCs in the context of local dynamics within the domains not considered, it is necessary to identify those residues with local flexibility or undergoing conformational exchange, so that they can be excluded from the structure and alignment tensor analysis.

Considerations regarding the order parameters As discussed in the introduction (section 1.4), the overall orientational distribution of a bond vector depends on both the molecular alignment and any movement of the bond within the molecular frame. The effect of small-amplitude, axially symmetric, local motions (that can be assumed to have no effect on the global molecular alignment) on the observed coupling can be accounted for by inclusion of an order parameter, S, in the pre-factor. Mathematically, this order parameter is equivalent to the Lipari-Szabo order parameter derived from analysis of relaxation data. However, while the Lipari-Szabo order parameter is sensitive only to motions on pico to nano second timescales, the RDC order parameter is sensitive to all motions up to a milli second timescale.

Analysis of the RDC data with respect to the structure and alignment tensor parameters would be more simple if the order parameter in the pre-factor can be assumed to be the same for all the analysed bond vectors, and hence discarded. The orientation of the extracted alignment tensor is unaffected, but there is a downscaling of the magnitudes of the principal components by a factor equal to the order parameter. To adopt this approach, only those RDCs measured from bond vectors that are rigid are retained during the analysis. In other words, only those bond vectors whose order parameter values are close to one are considered.

In order to allow the required filtering of the data according to order parameter, ^{15}N relaxation rates (R_2, R_1 and heteronuclear NOE) were recorded on isolated L7/L12 (data shown in Figure A.4 on page 214) and analysed according to the model-free formalism to

extract Lipari-Szabo order parameters [210].

Several previous studies have examined the dynamic properties of isolated L7/L12 proteins using NMR relaxation data [22, 34, 179]. The architecture of the protein, comprising two structured domains connected by a flexible linker, complicates the relaxation data analysis and in principle prevents a simple model-free analysis, which is based on the assumption of a single freely tumbling domain. In the earlier studies [34, 179], the analysis was restricted to extraction of a per-residue correlation time. The average correlation times for N- and C-terminal domians indicated that these domains do not tumble completely independently, despite the length of the flexible linker. A more recent study analysed the relaxation data using the reorientational eigenmode dynamics (iRED) approach [198], which was applied to a SAXS-optimised ensemble. The overall picture is that there is partial coupling between the rotational motion of the C-terminal and dimeric N-terminal domains, which is significantly stronger than that predicted assuming a completely flexible linker region. In addition, it was found from the iRED/SAXS analysis that there is an anti-correlation in the position of the two CTDs with respect to their proximity to the N-terminal dimer domain. Combined with the helical propensity of the residues in the hinge region, it has been proposed that the flexible hinge region exchanges between extended, random-coil and retracted/partial helix conformations, with the restriction that only one of the two linkers can form a helical conformation at any one time. On the ribosome, such a mechanism has been postulated to explain the observation that only two of the four L7/L12 CTDs appear to contribute to the observed NMR signal [34, 179].

All the relaxation studies mentioned above have considered a dimeric L7L12 species, with the dimer formed by interlocking of the helix hairpins of two NTDs. All spectra of L7/L12 acquired during this work showed markedly broader lines for residues in the NTD than for those in the CTD (Figures 2.5 and 2.12). An early explanation of this line-broadening proposed conformational exchange within the NTD [211], but a later study suggested that the slightly slower rotational reorientation of the NTD relative to the CTDs accounted for the broader linewidths [22, 179]. However, the diffusion data acquired on

fraction 2 of L7/L12 at varying concentration (Section 2.2.2.2) was strongly indicative of tetramerisation of the tight L7L12 dimer, most probably via the NTD. Such tetramerisation of the N-terminal dimer was not considered in the previous relaxation studies. The opportunity of further investigating this possible tetramerisation provided an additional motivation for undertaking a relaxation data analysis.

The measured relaxation rates (Figure A.4 on page 214) were analysed using the classical model-free formalism allowing for anisotropic rotational diffusion. As mentioned earlier, the model-free approach is not formally applicable to a multi-domain system with correlated interdomain motions, but in this case it is a very good approximation for the structured domains. In fact, the iRED analysis showed that the orientational eigenmodes can be re-expressed as anisotropic diffusion tensors for the C-terminal and N-terminal dimer domains. The overall rate of rotational diffusion can be characterised by a single isotropic rotational correlation time, which is related to the trace of the diffusion tensor.

Analysis of the relaxation data gave isotropic rotational correlation times of 10.8 ns and 9.1 ns for the N- and C-terminal domains, respectively. As reported previously, these correlation times are significantly longer than would be predicted from completely independent tumbling of the isolated domains (4.3 ns and 4.1 ns, respectively, derived using HYDROPRO), confirming the correlated motion of the domains. However, the extracted correlation times are also longer than those reported in the two previous relaxation studies (9.4 ns and 8.0 ns for the NTD, and 7.3 ns and 5.9 ns for the CTD, according to [22, 179]). In fact, all three sets of correlation times are different. The difference may be due to different degrees of tetramerisation of the NTDs, where different proportions of dimer and tetramer would lead to different apparent correlation times for different sample concentrations and conditions. Given that the motion of the CTD is not completely decoupled from the NTD, it is expected that a change in the average oligomerisation state of the NTD will also influence the rotational correlation time of the CTD. However, the effect should be rather small, especially as the effective correlation time of the CTD on the ribosome is only slightly longer than in isolated L7/L12 (Section 2.2.1.7). This indicates that, for the CTD at least,

the differences in rotational correlation time are more likely to be due to small changes in the conformational properties of the linker, which could alter its average flexibility, also then affecting the rotational correlation time of the NTD.

The order parameters for the N- and C-terminal domains are presented in Figure 2.17. For the subsequent analysis of the RDC data, residues with S^2<0.85 (S<0.92) were excluded.

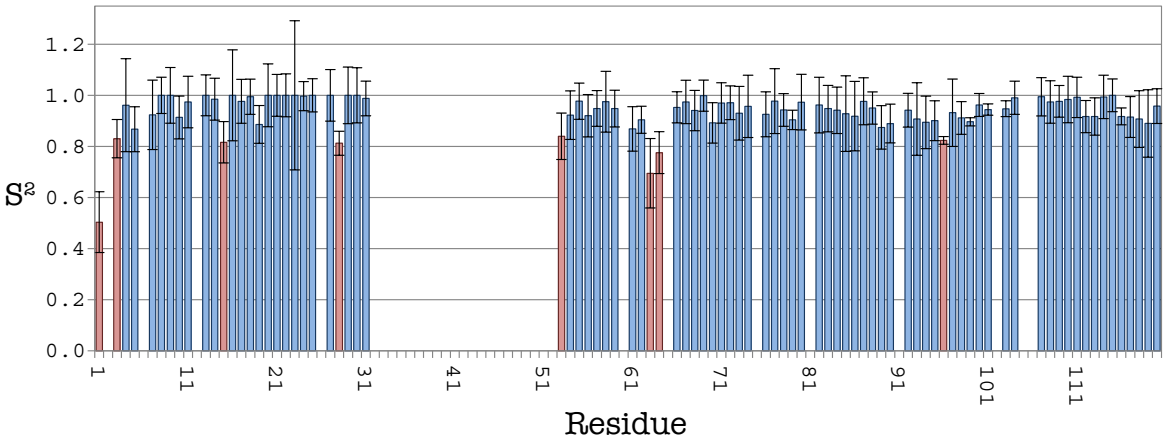

Figure 2.17: Plot of order parameters, S^2, of isolated L7/L12 protein, measured from fraction 2. The data points in red represent residues which are excluded for further RDC analysis.

Considerations regarding exchange contributions In addition to filtering the measured RDCs according to their order parameters, it is also necessary to identify those residues undergoing chemical or conformational exchange. In the context of RDCs, the influence of exchange appears through possible changes in both the local orientation of the internuclear vector and the overall alignment tensor. For conformational exchange, changes in the alignment tensor may be relatively small, in which case the changes in the local orientation are most important. In contrast, for oligomeric exchange, the local orientations are likely to be relatively unaffected, at least for the majority of residues, but the alignment tensors of the different oligomeric forms will be significantly different.

In the case of slow exchange on the chemical shift timescale, separate peaks are expected to be observed for those residues whose shift is affected by the exchange process. In general, the doublet components detected during the experiments to measure splittings will then also be in slow exchange, and separate sets of RDCs can be extracted for the two states. However,

residues in fast exchange on the chemical shift timescale will also be in fast exchange with respect to each line of the doublet, which means only one set of doublets will be observed. Therefore, it will not be possible to measure separate RDCs for the two states, and the observed RDC will be a population-weighted average. Analysis of such a system will be challenging since the observed RDCs will depend not only on the relative populations of the two states, but also on both the local structure and alignment tensors of those two states.

As discussed earlier, two fractions of isolated L7/L12 were observed after purification of recombinantly produced L7/L12. NMR, biochemical and biophysical analysis (Section 2.2.2.2) indicated that the two fractions differed in the region of the N-terminal two-helix domain. In particular, it was found for fraction 2 that the apparent hydrodynamic radius was dependent on the concentration, indicating protein-protein association, while the nature of the species in fraction 1 remains unknown. These observations suggested that detailed analysis of the N-terminal RDCs from fractions 1 and 2 of isolated L7/L12 would be intractable.

Estimation of exchange contributions to the transverse relaxation To provide further evidence of the underlying exchange process and also to identify possible conformational exchange in the CTD, a relaxation-dispersion experiment was recorded on fraction 2 of L7/L12.

This experiment is similar to the CPMG sequence commonly used to measure the ^{15}N transverse relaxation rate. However, rather than maintaining a constant CPMG frequency and varying the relaxation delay period, the dispersion experiment is based on a constant-time CPMG relaxation period for which the CPMG frequency is varied [212]. This yields a relaxation-dispersion profile, in which the effective relaxation rate is plotted as a function of CPMG frequency. The profile can be fitted to different models to extract parameters describing the underlying exchange process such as relative populations and exchange rates.

In the simplest implementation, a two-point experiment is recorded, using only the fastest and slowest accessible CPMG pulsing rates. An approximate estimate for the exchange contribution to the transverse relaxation rate (R_{ex}) is simply the difference between the

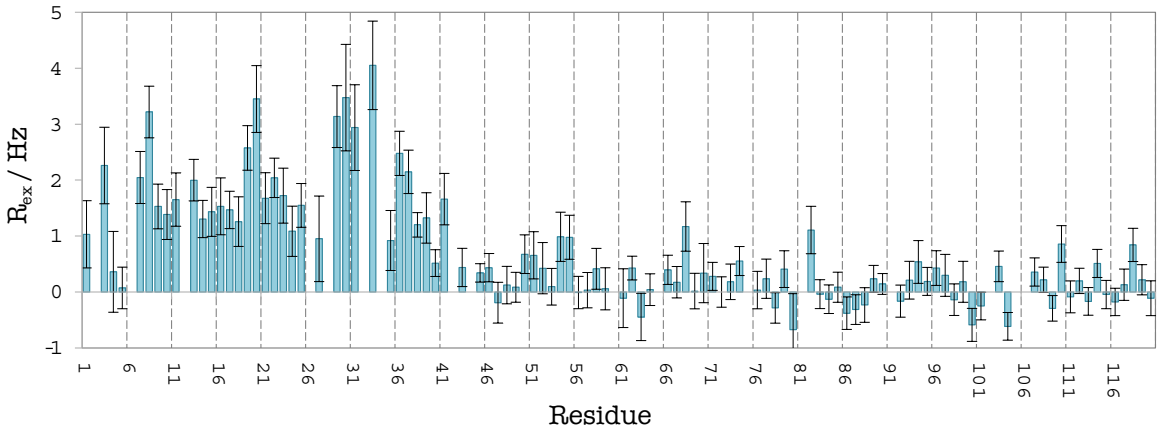

Figure 2.18: Plot of exchange contributions to the transverse relaxation rate, R_{ex}, of isolated L7/L12 protein, measured for fraction 2.

effective relaxation rates for the two CPMG frequencies. This estimation assumes that the effective relaxation rate at the slowest CPMG frequency is equal to the the free-precession transverse relaxation rate and that at the highest CPMG frequency is exchange-free. Even if this is not true, it provides a lower limit for R_{ex}.

The R_{ex} values for fraction 2 of L7/L12 estimated by this method are shown in Figure 2.18. The uncertainties are relatively large, being approximately of the order of the measured R_{ex} for the CTD. However, the majority of residues in the NTD showed measured R_{ex} values that were significantly greater than the estimated uncertainty, indicating that the RDC data from the NTD of the fraction 2 of L7/L12 will not be interpretable in terms of a single structure/alignment tensor.

This simple two-point experiment does not provide much information on the source of the exchange contributions, but supports the hypothesis that the observed concentration-dependent oligomerisation is mediated via the N-terminus, possibly by formation of a dimer of dimers. However, the possibility of additional conformational exchange contributions can not be excluded completely. Further separation of the two possible exchange processes would require undertaking dispersion experiments at varying protein concentrations.

Considerations regarding uncertainties in the data In order to be able to usefully interpret the measured RDCs, the error or uncertainty in their measurement must be significantly smaller than the range of observed values for a particular coupling type and

alignment strength. Where the RDCs are extracted by measuring the frequency difference between the two components of a doublet, the uncertainty depends on the accuracy of the peak-picking, which in turn depends on the signal-to-noise ratio and linewidth of the peaks.

An empirical relationship between the peak position uncertainty and the signal-to-noise ratio and peak linewidth was derived using a large set of simulated peaks with varying signal-to-noise ratios and linewidths. The relationship can be expressed:

$$\delta\nu = \frac{1}{2} \cdot \frac{LW}{SN} \tag{2.2}$$

where $\delta\nu$ is the uncertainty of the peak position in Hz, LW is the linewidth (full-width at half-height, in Hz) and SN is the signal-to-noise ratio, which is the ratio of the peak height to the RMS spectral noise.

The RDC is calculated from four peak positions:

$$D = 2 \cdot \left[(\nu_{\text{align,Hsqc}} - \nu_{\text{align,Trosy}}) - (\nu_{\text{isotropic,Hsqc}} - \nu_{\text{isotropic,Trosy}}) \right] \tag{2.3}$$

and hence the corresponding uncertainty is given by:

$$\delta D = 2 \cdot \sqrt{\delta\nu^2_{\text{align,Hsqc}} + \delta\nu^2_{\text{align,Trosy}} + \delta\nu^2_{\text{isotropic,Hsqc}} + \delta\nu^2_{\text{isotropic,Trosy}}} \tag{2.4}$$

which can be calculated by substituting Equation 2.2 for each of the $\delta\nu$ terms.

2.2.3.5 Analysis of the RDCs measured on isolated and ribosome-bound L7/L12

The measured RDCs from well-structured, non-exchanging residues in L7/L12 were analysed for the two fractions, fractions 1 and fraction 2, of isolated L7/L12 and ribosome-bound L7/L12. The analysis was conducted in two parts. In the first part, the measured RDCs were used in combination with the NMR structure of L7/L12 to extract the corresponding alignment tensors, using the singular value decomposition (SVD) method implemented in the program PALES [213]. In the second stage, the data were used for structure refinement of the

CTD NMR structure, confirming that the structure of the C-terminal domain of ribosome-bound L7/L12 is the same as in the isolated protein.

Analysis of RDCs from the N-terminal domain (NTD) of isolated L7/L12 Rigorous interpretation of the measured RDCs from the NTDs of the two fractions of isolated L7/L12 was prevented by the presence of probable dimer-tetramer exchange for fraction 2, and by the unknown architecture and peak assignments of the N-terminal species in fraction 1. Nevertheless, some qualitative observations for fraction 2 are possible.

For this fraction, the range of RDCs measured on the NTD was small compared to the CTD (Figure A.3 on page 213). This was most likely due to the overall negative charge on the surface of NTD, which limits the interaction of the domain with the negatively charged phage, leading to overall weak alignment.

Fitting the RDCs from fraction 2 to the NMR structure of the NTD yielded a Q factor of 0.28, with an RMSD of 0.4 Hz (Table 2.3). The Q factor is a magnitude-independent quality factor that describes the agreement between the measured (D_{meas}) and the back-calculated (D_{cal}) RDCs (Figure 2.19 A) according to the equation:

$$Q = \frac{\mathrm{RMS}(D_{meas} - D_{cal})}{\mathrm{RMS}(D_{meas})} = \frac{\mathrm{RMSD}}{\mathrm{RMS}(D_{meas})} \tag{2.5}$$

N	RMS_{err}/Hz	RMSD/Hz	Q	D_a^{NH}/Hz	δD_a^{NH}/Hz	R	δR
20	0.12	0.40	0.28	-1.2	1.1	0.46	0.14

Table 2.3: Table of quality and alignment tensor parameters after SVD fitting of measured N-H RDCs to the L7/L12 NTD structure for isolated L7/L12 in phage. N is the number RDCs measured, RMS_{err} is the root-mean-square measurement error (uncertainty), RMSD is the root-mean-square-deviation between the measured and back-calculated RDCs. Q is the quality factor, D_a^{NH} is the axial component of the alignment tensor scaled for the N-H coupling (Equation 1.18) and R is the rhombicity of the alignment tensor (Equation 1.19). δD_a^{NH} and δR are the uncertainties in D_a^{NH} and R, respectively.

The RMSD is significantly greater than the RMS uncertainty, indicating that the differences between the measured and back-calculated couplings are greater than can be accounted for by the measurement error. This difference could be partly due to structural

noise in the NMR structure, but is also likely to be due at least partly to the oligomeric exchange at the NTD.

Analysis of RDCs from the C-terminal domain (CTD) of isolated L7/L12 Both fraction 1 and fraction 2 of isolated L7/L12 showed a good correlation between the measured and back-calculated RDCs for the CTD (Figure 2.19 B & C), with Q factors of 0.21 and 0.19, respectively. However, the RMSDs between the measured and back-calculated RDCs are significantly larger than the experimental uncertainties (see Table 2.4), indicating that the differences are principally due to structural noise in the template structure, and suggesting that the data can be applied for structural refinement of the template structure. However, the low Q factors, which are of the order of those observed when fitting RDC data to high-resolution crystal structures, indicates that any structural deviations from the template are rather small.

	N	RMS_{err}	RMSD	Q	D_a^{NH}	δD_a^{NH}	R	δR
fraction1	38	0.59 Hz	4.6 Hz	0.21	-33.0 Hz	1.8 Hz	0.227	0.042
fraction2	50	0.17 Hz	2.9 Hz	0.19	-22.0 Hz	1.0 Hz	0.225	0.034

Table 2.4: Table of quality and alignment tensor parameters after SVD fitting of measured N-H RDCs to the L7/L12 CTD structure for isolated L7/L12 in phage.

The strong correlation between the measured RDCs for the CTDs of two fractions indicates that the CTD alignment tensors are very similar for these two species. The fact that the slope of the line of best fit is not equal to one simply reflects slightly different overall strengths of alignment, which is supported by examination of the D_a values for the derived alignment tensors (see Table 2.4). The D_a values, which reflect the strength of the alignment, are -33 Hz and -22 Hz for fraction 1 and fraction 2, respectively. These values are almost in exactly the same proportion as the measured deuterium splittings for the two phage-aligned samples (5.4 Hz and 3.7 Hz, respectively). The rhombicity of the two tensors, which is independent of the strength of alignment, is the same for both fractions, again suggesting that the alignment tensor is very similar for the two fractions

The similarity between the orientations of the two alignment tensors can be quantified

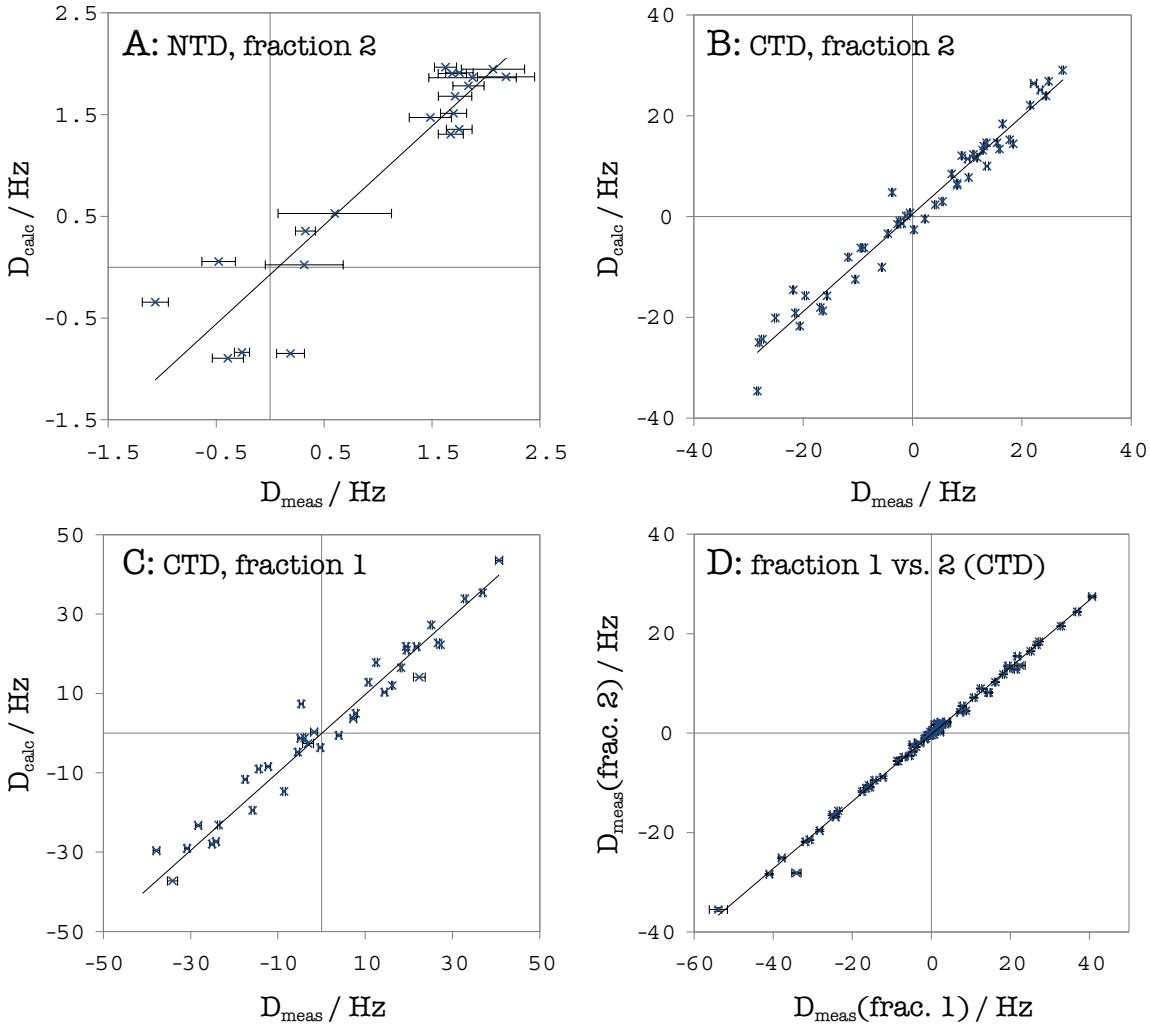

Figure 2.19: A-C: Correlation plots between the measured and back-calculated RDCs for the NTD and CTD of fraction 2 (A & B) and CTD of fraction 1 (C) for isolated L7L12. **D:** Correlation plot between the measured RDCs for the CTDs of fraction 1 and fraction 2.

by defining the angle between the tensors in 5-dimensional tensor space, which is calculated by taking the scalar product of the two tensors [135]. This calculation yields an inter-tensor angle of just 1.4 °. The significance of this small difference can be determined by visualising the orientation of the 1000 alignment tensors derived for each fraction during the Monte-Carlo error analysis. The spread of these orientations for each fraction can then be compared to the difference between the orientations for the different fractions. The results are shown in Figure 2.20 A, in which the orientations of the tensors are represented as Sanson-Flamsteed projections of the three orthogonal axes of the principal axis frames.

Together, these projections show that the orientations of the tensors for the CTDs are

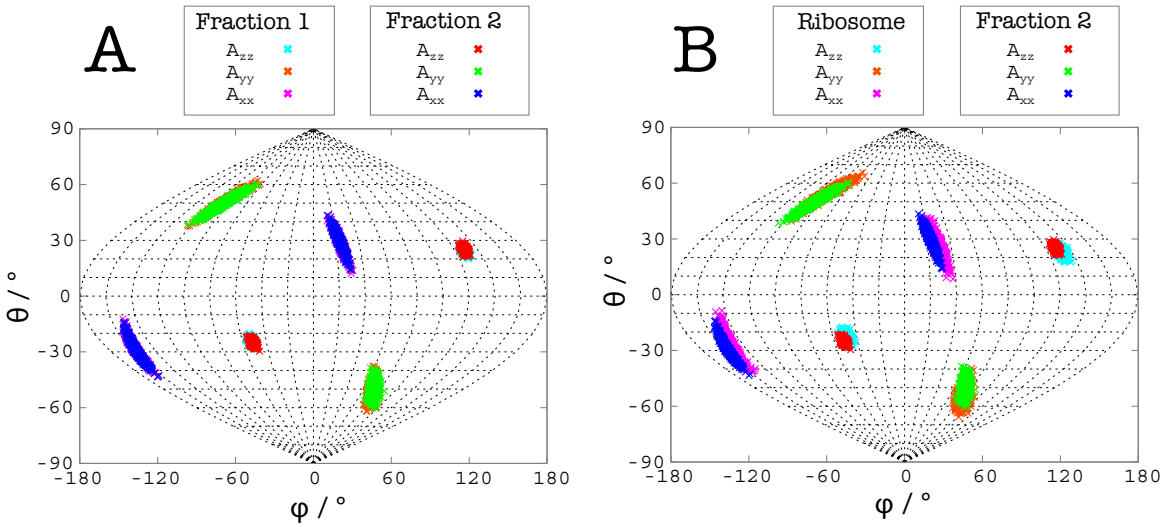

Figure 2.20: Sanson-Flamsteed projection showing orientations of the principal axes of the alignment tensor in the molecular PDB frame. **A:** Comparison of alignment tensor orientations derived for the CTDs of fraction 1 and fraction 2 of isolated L7/L12 in phage. **B:** Comparison of alignment tensor orientations derived for the CTD of ribosome-bound and isolated L7/L12 (fraction 2) in phage.

well-defined for both fraction 1 and fraction 2, with tight clustering of the axes. However, it is also clear that the spread in the orientation of the tensors, while relatively small, is greater than any difference between the two tensors. This is confirmed by the Monte-Carlo derived uncertainties in the tensor angle for the two tensors, which at 4.3 ° and 3.8 ° (for fraction 1 and fraction 2, respectively)are also larger than the angle between the tensors at 1.4 °.

These results show firstly that the two fractions of the isolated protein behave identically at their CTDs, further confirming that the difference between the two fractions lies at the NTD. Secondly and more interestingly, the indistinguishability of the C-terminal alignment tensors demonstrates that the differences at the N-terminus are not detectable at the C-terminus, at least via the phage-aligned dipolar couplings.

Analysis of RDCs for ribosome-bound L7/L12 (CTD) The correlation between the measured and back-calculated RDCs for ribosome-bound L7/L12 is shown in Figure 2.21 A. The correlation was not quite as good as for the isolated L7/L12 fractions, which is also reflected in the slightly higher Q factor of 0.3. This can be attributed to the relative uncertainties in the measured RDCs (RMS_{err}, Table 2.5), which are significantly higher for

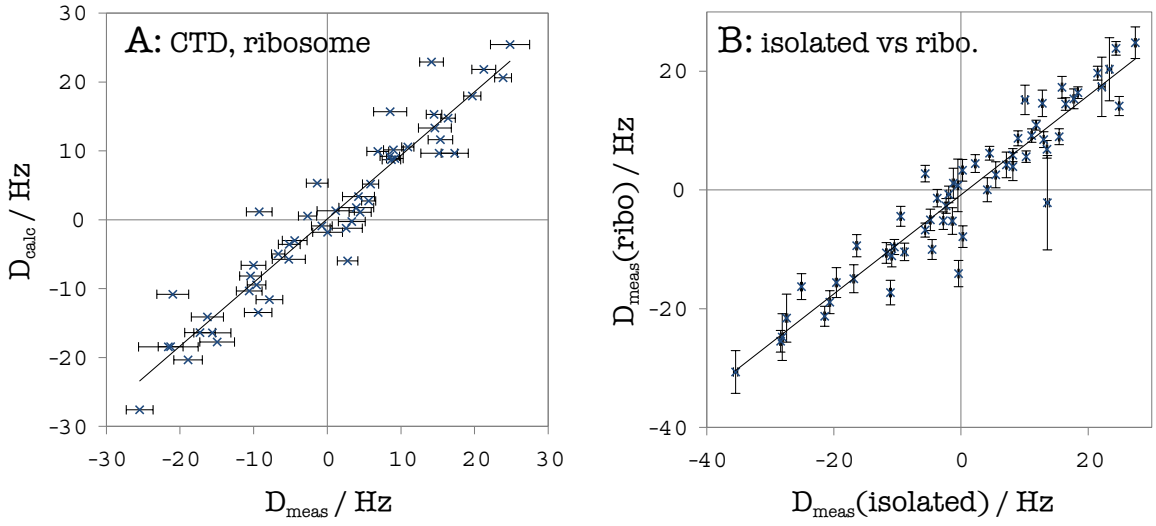

Figure 2.21: A: Correlation plot between the measured and back-calculated RDCs for the CTD of ribosome-bound L7/L12. **B:** Correlation plot between the measured RDCs for the CTDs of ribosome-bound and isolated L7/L12 (fraction 2).

the ribosome-bound L7/L12 due to the poorer signal-to-noise of the NMR spectra. The quality of the agreement between the measured and back-calculated RDCs is still sufficient to confirm that the structure of the ribosome-bound L7/L12 CTD is closely similar to that in the isolated protein. While the experimental errors are greater for the ribosome-bound L7/L12, the RMSD between the measured and back-calculated RDCs is approximately twice as large as the RMS uncertainty (Table 2.5), indicating that there is still scope for structural refinement using these data.

N	RMS_{err}	RMSD	Q	D_a^{NH}	δD_a^{NH}	R	δR
51	1.84 Hz	3.8 Hz	0.29	-18.2 Hz	1.2 Hz	0.299	0.051

Table 2.5: Table of quality and alignment tensor parameters after SVD fitting of measured N-H RDCs to L7/L12 CTD structure for ribosome-bound L7/L12 in phage.

The magnitude of the alignment tensor, D_a (Table 2.5), is slightly smaller for the ribosome data than for either of the two fractions of isolated L7/L12, while the deuterium splitting for the phage-aligned ribosome sample, at 12.3 Hz (Figure 2.14), was larger than for the isolated L7/L12 samples (5.4 Hz and 3.7 Hz). The higher phage concentration in the ribosome sample should lead to stronger alignment of the L7/L12 CTD than in the isolated L7/L12 samples, however, the linewidth of the deuterium signal was observed to

be significantly broader for the ribosome sample. This is an indication that the presence of the ribosome is possibly interfering with the homogeneity of the liquid crystal medium formed by alignment of the phage particles in the external field.

The orientation of the alignment tensor for the CTD of ribosome-bound L7/L12 can be compared with that for fraction 2 of isolated L7/L12 by overlaying the Sanson-Flamsteed projections of the principal axes orientations from the Monte-Carlo simulations for the two data-sets (Figure 2.20 B). The difference between the alignment tensor orientations for the ribosome-bound and isolated L7/L12 appears greater than that between the tensor orientations for the two fractions of isolated L7/L12 (Figure 2.20). However, the size of the differences between the tensor orientations for ribosome-bound and isolated L7/L12 is similar to the spread of each of the orientations within the set of repeated Monte-Carlo runs. Quantitatively, the angle between the best-fit alignment tensors was calculated to be 7.2 °, while the uncertainties in the angles of tensors were derived from the Monte-Carlo simulations as 5.3 ° and 3.8 ° for the ribosome-bound and fraction 2 of isolated L7/L12, respectively.

The rhombicity of the alignment tensor derived for the CTD of ribosome-bound L7/L12 is also slightly different to the rhombicities for the fractions of the isolated protein (R = 0.30 for the ribosome *vs* R = 0.23 for isolated L7/L12), but again the uncertainties in the rhombicities (0.034 and 0.051) are of similar magnitude as the difference between the fitted values. The application of a simple t-test to assess the significance of the differences between the angles and rhombicities of the alignment tensors indicates that the p-value for both parameters is approximately 25%, which is too high for rejection of the null hypothesis that there is no difference between the tensors. Therefore, while there appears to be a small difference in the phage-induced alignment tensors of the CTDs of isolated and ribosome-bound L7/L12, this difference cannot be asserted to be significant given the current uncertainties. However, it should be mentioned that the Monte-Carlo error analysis is not straightforward, as it needs to account for both the errors in the measured RDCs and possible errors in the template structure used to derive the alignment tensor (see Material and

methods section 5.4.4 for more details on the Monte Carlo-derived uncertainties).

2.2.3.6 Structural refinement of the CTD of ribosome-bound and isolated L7/L12

The alignment tensor analysis of the RDC data measured on isolated and ribosome-bound L7/L12 described above used the measured RDCs in combination with a template structure to derive the associated alignment tensor. The assumption for this approach is that the template structure is a good model for the 'true' structure. The agreement between the measured data values and those back-calculated using the template structure and fitted alignment tensor depends on both the measurement error and on any differences between the template and 'true' structures (structural noise). A root-mean-square deviation between the measured and back-calculated RDCs that is larger than the measurement error indicates that the structural noise is greater than the measurement noise, and hence that the experimental data can be used to refine the template structure.

The refinement is typically implemented in a simulating-annealing molecular-dynamics simulation by introducing an additional term to the force-field whose energy depends on the difference between the experimental RDC and that back-calculated from the structure. However, the power of RDC-based structural refinement relies on the simultaneous incorporation of multiple RDC-based restraints based on different types of internuclear vectors and/or measurements in multiple alignment media. This requirement for multiple restraint-types is partly due to the degeneracy in allowed bond-vector orientations consistent with a measured RDC for a single internuclear vector. This degeneracy is significantly reduced both by measuring couplings for multiple bond vectors that have known relative orientations within a rigid molecular fragment (such as the peptide plane) and by measuring the same couplings in different alignment media (see section 1.4.2).

For the CTD of L7/L12, only N-H couplings in phage-aligned samples were measured to date, limiting their application for RDC-based structural refinement. However, they can be applied within a restricted structural refinement scheme in which only the relative orientations of the secondary structure elements are allowed to change [214]. The

	N	RMS_{err}	RMSD	Q	D_a^{NH}	δD_a^{NH}	R	δR
isolated (before)	39	0.16 Hz	2.5 Hz	0.16	-21.9 Hz	0.9 Hz	0.230	0.028
isolated (after)	39	0.16 Hz	1.6 Hz	0.11	-21.4 Hz	0.7 Hz	0.208	0.023
ribosome (before)	42	1.87 Hz	3.8 Hz	0.30	-18.3 Hz	1.4 Hz	0.305	0.053
ribosome (after)	42	1.87 Hz	3.1 Hz	0.25	-17.8 Hz	1.1 Hz	0.284	0.045

Table 2.6: Table showing comparison of quality and alignment tensor parameters before and after structural refinement of isolated and ribosome-bound L7/L12 CTD.

local backbone structure remains essentially unchanged, such that the secondary structure elements can be considered as rigid molecular fragments. Since several N-H couplings are available for each alpha-helix or beta-strand, the orientation of the helices/strands are well defined by the RDC data.

The template structure for the CTD of L7/L12 was refined using the restricted approach described above, implemented in Xplor-NIH [215]. The local structure of the secondary structure elements was fixed using non-crystallographic symmetry (NCS) restraints, and only those RDCs corresponding to rigid N-H vectors within the secondary structure elements were included as restraints. The values for the magnitude and rhombicity of the alignment tensor for back-calculation of the RDCs were taken as those of the best-fit alignment tensor derived using the template structure. The orientation of the alignment tensor is allowed to float during the refinement. The RDC restraints were implemented as soft-square-well susceptibility anisotropy (SANI) restraints, with the width of the well adjusted according to the experimental uncertainty of each measured RDC.

The structure refinement was performed using the measured RDCs for both isolated fraction 2 and ribosome-bound L7/L12. The parameters for the SVD fitting of the secondary-structure RDCs using the template and refined structures for the two data-sets are shown in Table 2.6. For both data-sets, there is a small but significant reduction in both the Q factors ($0.3 \rightarrow 0.25$ and $0.16 \rightarrow 0.11$ for ribosome-bound and isolated L7/L12, respectively) and the RMSDs between the measured and fitted RDCs (3.79 Hz → 3.11 Hz and 2.46 Hz → 1.62 Hz), indicating that the refined structure is closer to the 'true' structure than the initial template structure. However, the RMSD values are still larger than the experimental uncertainties, which demonstrates that there remains a significant degree of noise within the local structure

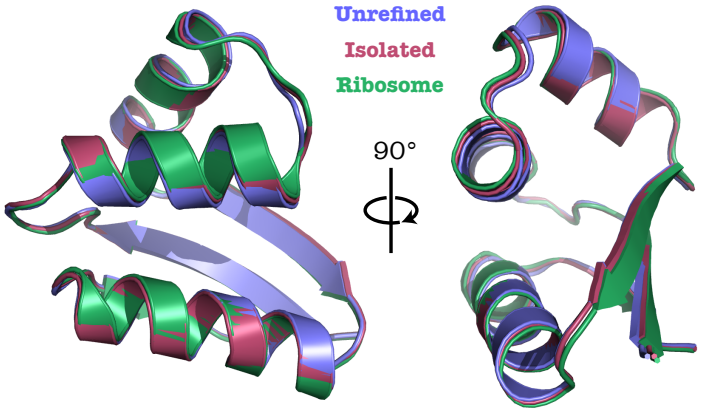

Figure 2.22: Overlays of the template NMR structure and the refined structures of the CTD for ribosome-bound and isolated L7/L12 (fraction 2).

of the refined structures that could not be corrected using this restricted refinement protocol.

Overlays of the refined structures for the CTDs of ribosome-bound and isolated L7/L12 with the template structure and with each other are shown in Figure 2.22. It is clear that the differences between the refined and template structures are rather small, as was expected given the low initial Q-factors obtained from fitting the measured data to the template structure. The pairwise backbone RMSDs to the template structure are 0.52 Å and 0.35 Å for the ribosome-bound and isolated refined structures, respectively. The backbone RMSD between the two refined structures is even smaller, at 0.28 Å, further highlighting that the CTD structure of L7/L12 is the same on the ribosome as in the isolated protein.

The alignment tensor analysis for the CTDs described in Section 2.2.3.5 was repeated using the refined structures. However, no significant differences from the analysis using the initial template structure were apparent. Both the rhombicity and the angle of the two alignment tensors changed slightly, but the difference between the CTD alignment tensors for the ribosome-bound and isolated L7/L12 derived using the respective refined structures, while maintained, remained insignificant relative to the estimated uncertainties.

2.2.3.7 Discussion

The principal motivation behind measuring RDCs for L7/L12 on the ribosome was to demonstrate the feasibility of measuring RDCs on the ribosomal particle and to develop

the required methodology. However, the behaviour of the L7/L12 protein on the ribosome is also of interest. The L7/L12 protein is important for protein synthesis, acting as a recruiter for the translation factors IF2, EF-Tu, EF-G and RF3. Despite its importance, obtaining structural information on ribosome-bound L7/L12 has proved difficult. The flexibility of the CTD makes it invisible by the crystallographic methods that have been used to determine the structure of the majority of the ribosomal particle. In contrast, the flexibility allows the CTD to be observed by solution-state NMR, while the signals from the structured ribosomal core are too broad to be detectable. Hence, solution NMR is the most promising technique for probing the structure and dynamics of ribosome-bound L7/L12.

Prior to measurement of RDCs from ribosome-bound L7/L12, data were acquired for two structural forms of the isolated L7/L12 protein. These two forms exhibited different hydrodynamic properties, eluting at different volumes under gel-filtration and having different translational diffusion coefficients. The NMR spectra indicated that the two fractions differ at their N-termini. Fraction 2 appears to undergo a dimer-tetramer oligomerisation, mediated via the NTD, while fraction 1 has a different but unknown N-terminal architecture. The availability of these two fractions, which are the same at their C-termini but different at their N-termini, allowed consideration of the conformational independence of the domains.

Previous relaxation studies on isolated L7/L12 have shown that while the linker is highly flexible, the orientational motion of the two domains is not completely independent. In principle, the 20-residue linker should be long enough to make the relative orientationals distribution of the N- and C-terminal domains completely isotropic. However, it has been proposed that each linker within the dimer is in exchange between a random coil conformation and a conformation with partial helical character in the N-terminal half of the linker sequence. In addition, it has been noted that such a part-helical conformation can only be adopted by one of the pair of linkers at any one time, leading to an anti-correlation between the spatial positions of the two CTDs. The incomplete flexibility of the linker means that the orientation of one domain in a global frame should depend partly on the orientation

of the other domain in the same frame. For RDCs, this means that the alignment tensors of each domain are not only determined by the domains themselves, but are also influenced by the alignment of the other domain.

This hypothesis can be tested by considering the alignment tensors derived for the CTDs of the two fractions of isolated L7/L12. It was found that these tensors are identical within the experimental error, which, since the N-termini are different, suggests that the alignment of the CTD is independent of the alignment of the NTD. One explanation for this observation, which contradicts previous conclusions, would be that the two domains are conformationally independent. However, it was also found that the measured RDCs for the N-terminal domain of the fraction 2 were much smaller than those for the CTD, indicating the alignment of the NTD was relatively weak. Although the structure of the NTD in the fraction 1 is unknown, its alignment is also likely to be weak. Hence, the difference between the alignment tensors of the two NTDs is likely to be small, and hence there will be little difference in their influence on the alignments of the CTDs. In addition, the relative orientations of the two domains in phage-aligned samples will be determined mainly by the interaction of each domain with the phage medium. In other words, any weak preferred relative orientation of the domains that might exist under isotropic conditions could be swamped by the effect of the aligning force on the individual domains from their interaction with the phage. An alternative RDC-based approach to identifying restricted flexibility within the linker would be to use lanthanide-induced alignment by tagging either the N- or C-terminus. In this system, any alignment of the other, non-tagged, domain must come from propagation of the alignment of the tagged domain via the linker. In the case that the two domains were completely orientationally independent, the non-tagged domain would show no net alignment. Conversely, if non-zero alignment were observed for the non-tagged domain, then the two derived alignment tensors could be used to determine the preferred relative orientation of the two domains.

Following the RDC measurements on the isolated protein, data were acquired for ribosome-bound L7/L12. This set of RDCs represents the first atomic-resolution structural information for L7/L12 on the ribosome. The data confirmed that the structure of the CTD

of L7/L12 on the ribosome is the same as that of the isolated protein. In principle, the RDC data should also contain information on the preferred orientation of the L7/L12 CTDs relative to the ribosomal particle. However, extracting such information relies on being able to simultaneously measure some RDCs from the structured core of the ribosome, which was not attempted during this work and may not be possible. In addition, such attempts would be best made using lanthanide-induced alignment of the CTD, so that any small preference for a defined relative orientation of the L7/L12 and ribosome would not be affected by alignment forces acting on both the ribosome and L7.

The current data can still be used to assess the flexibility of the linker by comparing the alignment tensors for the CTD of isolated and ribosome-bound L7/L12. As described above for the two fractions of isolated L7/L12, differences in the alignment tensors of the NTD should result in smaller differences in the tensors for the corresponding CTDs, provided the linker is not completely flexible. In the case of the isolated protein, any difference between the alignment of the NTDs is probably small, so that possible differences in the tensors of the CTDs were undetectable. However, the alignment tensors of isolated and ribosome-bound NTD should be more different, which might lead to detectable differences in the alignment tensors of the corresponding CTDs, providing evidence for restricted flexibility of the linker. The derived alignment tensors for the ribosome-bound and isolated L7/L12 CTDs showed differences that were larger than those between the CTD tensors for the two fractions of isolated L7. Although these differences were judged not to be highly significant when compared to the Monte-Carlo derived uncertainties, the probability that the observed differences are due to chance was sufficiently low (~25%) to allow speculation that the difference is real. This conclusion would support previous findings regarding the restricted flexibility of the linker.

The above discussion has not considered the possible effects on the RDC data arising from the dimeric state of the isolated L7/L12 protein and additionally for the ribosome, the presence of two copies of the dimer. All the molecular species for which RDC data were acquired during this work showed just a single splitting for each N-H pair. Consider first

the isolated protein. If the structure of the dimer is asymmetric with respect to the CTD, as for example would be the case if one linker is contracted while the other is extended, then each N-H signal would show two splittings. Only one splitting is observed, which indicates that the two linkers are exchanging between the contracted and extended states on a timescale that is faster than any difference between the RDCs of the two states. On the ribosome, there are two L7/L12 dimers, but only two CTDs appear to contribute to the observable NMR signal. The current picture is that each dimer exists in a configuration where one linker is contracted and the other extended, such that two CTDs from different dimers are extended away from the ribosome, while the other two are close to ribosome. That the NMR signal appears to arise from just two copies of the CTD suggests that two CTDs attached to the linker in its contracted state are NMR-invisible. It follows that any exchange between the contracted and extended linker conformations must be slow relative to timescale of the NMR experiment, or all of the L7/L12 signal would disappear. This is different to the isolated protein, where the linkers seem to be exchanging between contracted and extended conformations on a timescale fast enough to average any differences in RDCs. Returning to the ribosome, the picture of two L7/L12 dimers, each with one extended and one contracted linker, suggests that two sets of RDCs may be observable. The dimers have different positions on the underlying L10 protein, and hence the two extended CTDs, one attached to each dimer, would be expected to have different preferred relative orientations to the ribosome. These different relative orientations should be reflected in different RDCs for each CTD assuming that the alignment tensor of each domain is partly determined by the alignment tensor of the ribosome itself. During this work, only one set of RDCs was observed. Two explanations can be proposed from this observation. Since the influence of the ribosome alignment tensor on the alignment tensor of the CTDs is expected to be small, particularly in phage-aligned samples, any small differences in the RDCs for the two observable CTDs may be undetectable within the precision of the experiment. Alternatively (or additionally), it is possible that the L7/L12 dimers are exchanging positions on the L10 molecule, on a timescale fast relative to any difference in the RDCs. Since any differences

in the RDCs are expected to be small, the longest timescale consistent with this explanation would be quite slow.

2.2.4 Analysis of C^{α}-H^{α} RDCs measured in phage

In order to build a better representation of the structural orientation of both ribosome-bound and isolated L7/L12, an approach examining the C^{α}-H^{α} RDCs was attempted.

2.2.4.1 Experimental approaches to measure C^{α}-H^{α} RDCs

Together with the one-bond ^{15}N-^{1}H coupling, the one-bond $^{13}C^{\alpha}$-$^{1}H^{\alpha}$ RDC is one of the most commonly measured RDCs within the protein backbone. One advantage of measuring C^{α}-H^{α} RDCs, which is important for ribosomes and RNCs, is that they are large in magnitude, which reduces the measurement precision required. The strength of the C^{α}-H^{α} dipolar coupling interaction is approximately twice that of the N-H interaction, so that RDCs of several tens of Hz are easily achievable with only weak alignment.

For structural refinement, their usefulness in combination with N-H couplings is not quite the same as for other couplings within the peptide plane. This is because the relative orientation of the C^{α}-H^{α} bond and the N-H bond of the same and successive residues is not fixed by the local backbone structure, but depends on the backbone angles ϕ and ψ. In other words, the relative orientations of these two pairs of internuclear vectors define the backbone angles ϕ and ψ. However, while the relative orientations of the internuclear vectors are partly restricted, they are not fixed, and hence only partly reduces the degeneracy of each of the orientations of the individual vectors relative to the alignment tensor. Therefore, their orientation relative to each other and relative to the alignment tensor is not uniquely defined by the pairs of measured RDCs. In contrast, the relative orientation of two vectors in the peptide plane is fixed by the molecular geometry such that their joint orientation relative to the alignment tensor is well defined.

Despite their large magnitude, it is more difficult to measure the C^{α}-H^{α} RDCs than the backbone N-H couplings experimentally. The principal difficulty arises from the carbon-

carbon couplings that are active in uniformly ^{13}C-labelled samples. The C^{α} nucleus exhibits J couplings to both C' and C^{β} nuclei (except in glycine), with magnitudes that are significant relative to the C^{α}-H^{α} J coupling. While the C'-C^{α} J coupling can be easily decoupled due to the large chemical shift separation of the C^{α} and C' nuclei, this is more difficult for the C^{α}-C^{β} interaction. Therefore, the C^{α}-C^{β} coupling interferes with the measurement of the C^{α}-H^{α} coupling as a result of their simultaneous evolution.

Three approaches towards handling the problem of C^{α}-C^{β} coupling evolution have been considered. First, the time period for evolution of the C^{α}-H^{α} coupling can be restricted so that evolution of the C^{α}-C^{β} coupling is limited. This is possible because the C^{α}-C^{β} J coupling is approximately four-times smaller than the C^{α}-H^{α} coupling. However, the resulting maximum evolution time ($< 1/3J_{C^{\alpha}C^{\beta}} \sim$10 ms) possible without resolving the C^{α}-C^{β} coupling is too short to completely resolve the C^{α}-H^{α} coupling, particularly as the unresolved C^{α}-C^{β} coupling will contribute to the effective linewidth. Hence, it will be necessary to record the spectra in an IPAP fashion. However, even then, the linewidths of the separated multiplet components will be broad, which compromises the precision of the measurement.

The second approach is to implement the C^{α} shift and C^{α}-H^{α} coupling evolution in a constant-time fashion. Since the C^{α}-C^{β} coupling is active over the complete constant-time period for all points, the resulting signal is not modulated by the C^{α}-C^{β} coupling. The length of the constant-time period is set to $(1/J_{C^{\alpha}C^{\beta}}) \sim$28 ms to give maximum transfer of signal. This approach has the advantage that it yields narrow lines in the C^{α} dimension, but it also suffers from much lower sensitivity due to the fast C^{α} transverse relaxation.

The third approach is to implement band-selective C^{β} decoupling during the C^{α} shift and C^{α}-H^{α} coupling evolution times, although this is challenging due to the overlapping chemical shift ranges of the C^{α} and C^{β} nuclei for some residue types. In particular, serine, threonine and leucine C^{β} nuclei resonate at shifts close to or overlapping with the C^{α} resonances.

Each of the above approaches can be implemented into different types of pulse sequence which contain an appropriate indirect C^{α}-evolution period. These sequences fall into

two categories: 2D ^{13}C HSQC-based sequences and 2D/3D HN-detected triple-resonance sequences. Constant-time ^{13}C HSQCs have been used to measure C^{α}-H^{α} couplings in a J-modulated fashion [156, 216]. A real-time IPAP-^{13}C-HSQC sequence was used to measure C^{α}-H^{α} couplings for selectively 13C^{α}-labelled samples, in which no other carbon sites were ^{13}C-labelled, thus there were no homonuclear carbon-carbon couplings [217]. A 3D IPAP-HNCA experiment was developed to allow simultaneous measurement of ^{1}D (C^{α}-H^{α}) and ^{2}D (N-H^{α}), using a short $^{13}C^{\alpha}$ evolution time to restrict C^{α}-C^{β} coupling evolution [218]. A similar approach was adopted for a 3D HNCO(J-CA) sequence, where the ^{13}C' chemical shift and C^{α}-H^{α} coupling evolution are implemented in a simultaneous accordion fashion [219]. Several sequences have been proposed based on the (HA)CA(CO)NH sequence, either via quantitative J-modulation [220, 221] or via measurement in the frequency domain with a 28-ms constant-time period for C^{α} shift/C^{α}-H^{α} coupling evolution [222, 223].

For considering which method is most appropriate for measuring C^{α}-H^{α} RDCs on the ribosome and RNCs, sensitivity of the sequence is of most importance. In general, the triple-resonance sequences, whether recorded in 2D or 3D mode, will be less sensitive than those based on the 2D ^{13}C HSQC sequence as a result of the multiple transfer steps in the triple-resonance pulse sequences. From this perspective, an experiment based on the 2D ^{13}C HSQC sequence may be most suitable for ribosomes/RNCs.

Even without the problems caused by C^{α}-C^{β} coupling, the ^{13}C HSQC experiment is not as good for measuring C^{α}-H^{α} couplings as is the ^{15}N HSQC experiment for measuring N-H couplings. In general, the C^{α}-H^{α} region of the ^{13}C HSQC spectrum is more overlapped than the ^{15}N HSQC spectrum. Additionally, the water signal falls in the middle of the H^{α} chemical shift range, which can reduce the quality of ^{13}C HSQC spectra acquired on samples in H_2O.

The resolution of the ^{13}C HSQC spectrum can be maximised by recording the ^{13}C chemical shift evolution with a constant-time period. However, the loss of sensitivity due to the long constant-time period required to refocus the C^{α}-C^{β} coupling (28 ms) and the fast C^{α} transverse relaxation will be too much for the low-concentration ribosome and RNC

samples. However, recording the ^{13}C HSQC spectrum with real-time ^{13}C shift evolution usually limits the ^{13}C acquisition time to less than 10 ms. The resulting broad linewidths will lead to extensive peak overlap in the C^{α}-H^{α} region, preventing accurate measurement of peak positions for a large fraction of the residues. Acquiring the ^{13}C HSQC spectrum with both sufficient sensitivity for ribosomal samples, and with the required resolution for extracting splittings, requires real-time ^{13}C shift evolution in combination with selective C^{β} decoupling. However, even with the extended acquisition times possible with by C^{β}-decoupling, the resolution is limited by the C^{α} transverse relaxation rate, so that the resulting spectrum will still suffer from peak overlap to some extent.

The 3D triple-resonance experiments are intrinsically less sensitive than the real-time ^{13}C HSQC due to the additional delays required for magnetisation transfer. However, they have the advantage over the ^{13}C HSQC approach that there will be minimal peak overlap, allowing measurements of splittings for all residues. The experiments can be constructed based on two sequences, HN(CO)CA and (HA)CA(CO)NH. The HN(CO)CA spectrum is usually acquired with real-time ^{13}C shift evolution (without C^{β} decoupling), and hence would need to be implemented with IPAP to separate the otherwise overlapping doublet components. The (HA)CA(CO)NH spectrum is acquired with constant-time ^{13}C shift evolution to allow simultaneous evolution of the C^{α}-C' coupling ready for the subsequent transfer step. As with the ^{13}C HSQC, this constant-time period can be adjusted to refocus the C^{α}-C^{β} coupling (28 ms), but again this will result in significant and unaffordable sensitivity losses. Hence, the constant-time period is tuned to the first maximum in the relevant transfer function, which occurs at ~6.4 ms. This is even shorter than for the HN(CO)CA, but there will no contribution of the unresolved C^{α}-C^{β} couplings to the linewidth in the (HA)CA(CO)NH spectrum, and the constant-time evolution allows safe extension of the time-domain signal to twice its original size using linear prediction. However, the (HA)CA(CO)NH experiment will also need to be recorded with IPAP. The sensitivities of the HN(CO)CA and (HA)CA(CO)NH sequences are expected to similar. The HN(CO)CA sequence can also be implemented in a BEST (band-selective excitation short-transient)

fashion, which could yield small sensitivity benefits [224].

Three sequences were tested for measuring C^{α}-H^{α} splittings on the ribosome: a 2D IPAP-^{13}C-HSQC sequence with real-time $^{13}C^{\alpha}$ evolution and selective C^{β}-decoupling, an IPAP-BEST-HN(CO)CA sequence and an IPAP-(HA)CA(CO)NH sequence.

$^{13}C^{\alpha}$-HSQC with selective C^{β} decoupling The basis of the IPAP-^{13}C HSQC sequence was a C^{α}-selective ^{13}C HSQC sequence with band-selective C^{β} decoupling. This sequence was developed from a standard gradient-selected, sensitivity-enhanced ^{13}C HSQC sequence with C' decoupling during the ^{13}C evolution period (Dr John Kirkpatrick, UCL). C^{α}-selectivity was achieved using a C^{α}-selective inversion pulse in the first INEPT. In addition, the refocusing pulse at the end of the ^{13}C evolution period, which is required to allow incorporation of the coherence-order selection gradient, was implemented as a C^{α}-selective pulse. C' and band-selective C^{β} decoupling were implemented using a doubly-cosine-modulated pulse, adjusted to give inversion bands for the C' nuclei, the downfield threonine C^{β} nuclei and the upfield C^{β} nuclei of all other residues (except serine and glycine). The C^{α}-selective pulses were adjusted to cover all C^{α} resonances except for glycines, which are therefore not present in the resulting spectrum. This allows for better coverage of the C^{β} resonances by the C'/C^{β} inversion pulses. Further details are given in the Materials & Methods, Section 5.3.

The selective $^{13}C^{\alpha}$-HSQC of isolated L7/L12 in Tico buffer is compared with the H^{α}/C^{α} region of the corresponding CT-^{13}C-HSQC in Figure 2.23 (the assignment of the spectrum is discussed in Section 2.2.4.3). In the $^{13}C^{\alpha}$-HSQC spectrum, all the glycine C^{α} and threonine C^{β} resonances are missing, but the remainder of the C^{α} peaks appear with uniform intensities. In particular, the downfield valine C^{α} peaks are present with good intensity. The only other difference is that the serine C^{α} and C^{β} peaks appear split due to their mutual coupling. Comparison of the signal intensities over a sub-set of well-resolved, dispersed peaks indicates that the $^{13}C^{\alpha}$ HSQC sequence is approximately 2.5-3 times more sensitive than the CT-^{13}C-HSQC sequence.

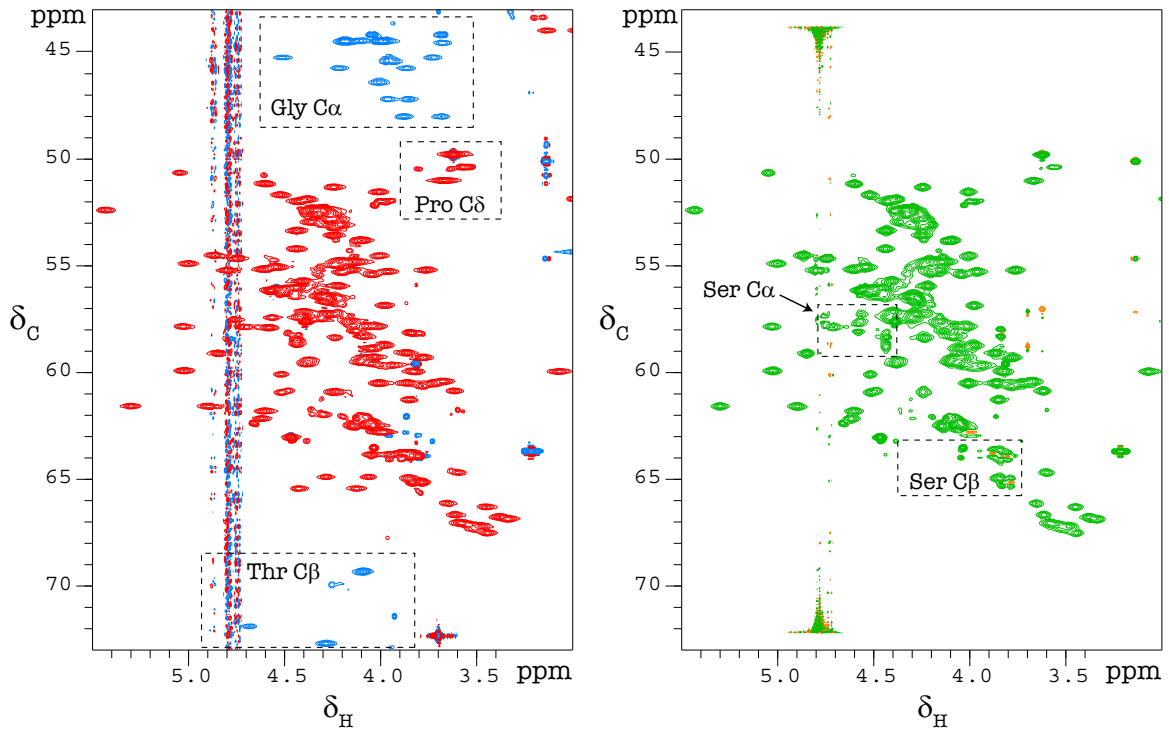

Figure 2.23: Comparison of CT-^{13}C-HSQC (red) and selective $^{13}C^{\alpha}$-HSQC (green) spectra. In the $^{13}C^{\alpha}$-HSQC spectrum, glycine C^{α} and threonine C^{β} resonances are missing, and the serine C^{α} and C^{β} peaks appear split. Both spectra were recorded on 250-µM isolated L7/L12 (fraction 2) in Tico buffer at 25 °C.

2.2.4.2 Comparison of 2D C^{α}-HSQC and 3D triple-resonance approaches

While being relatively sensitive, the 2D C^{α}-HSQC spectrum still suffers from peak overlap, which prevents accurate peak picking for the overlapped resonances. For this reason, the 3D triple-resonance approaches, based on the HN(CO)CA or (HA)CA(CO)NH sequences are preferred if sensitivity is not a limiting factor. To determine the feasibility of using the 3D triple-resonance approach on ribosome samples, a sample of isolated L7/L12 was prepared at 20 µM (in phosphate buffer), which is equivalent to the effective concentration of L7/L12 in 10 µM ribosome samples. A 2D in-phase H^{N}-C^{α} spectrum was recorded using the IPAP-BEST-HN(CO)CA sequence and compared to the 2D in-phase H^{α}-C^{α} spectrum recorded using the IPAP-C^{α}-HSQC sequence (Figure 2.24). Both spectra were recorded for 6 hours. The IPAP-C^{α}-HSQC sequence is clearly more sensitive than the IPAP-BEST-HN(CO)CA sequence. In the in-phase IPAP-C^{α}-HSQC spectrum, nearly all the peaks from the C^{α}-HSQC spectrum are present, whereas only a small fraction are visible above the noise level in the

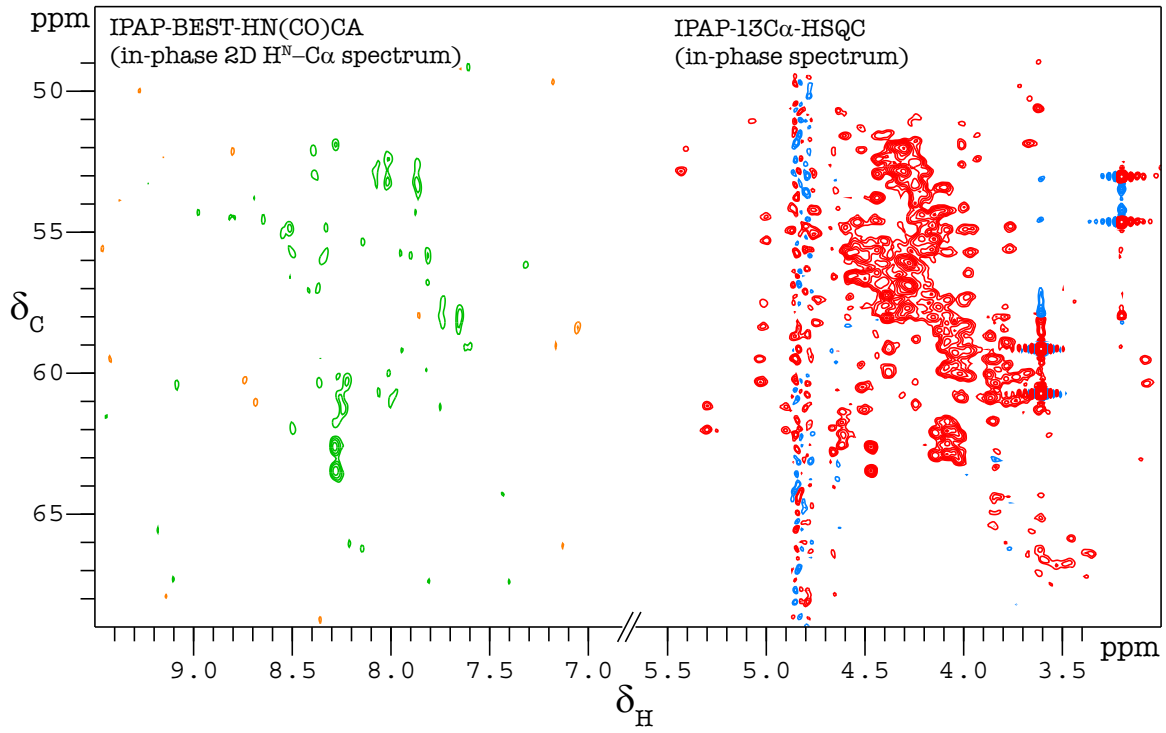

Figure 2.24: Comparison of the in-phase 2D H^N-C^α spectrum recorded with the IPAP-BEST-HN(CO)CA sequence (green) and the in-phase 2D IPAP-C^α-HSQC spectrum (red). Both spectra were recorded on 20-µM isolated L7/L12 (fraction 2) in phosphate buffer at 25 °C for 6 hours and contoured at the same level (3.5× RMS noise.

2D H^N-C^α in-phase spectrum.

For a ribosome sample, the data can be recorded for 48-72 hours, leading to a signal-to-noise improvement of 2.8-3.5-fold for the 2D H^N-C^α plane. However, the data would also then be acquired in 3D mode, reducing the signal-to-noise (S/N) due to the additional dimension. In order to determine more accurately if the HN(CO)CA approach would yield sufficient sensitivity, and to compare the HN(CO)CA and (HA)CA(CO)NH sequences, 3D in-phase/anti-phase spectra were recorded on the same 20-µM sample using the IPAP-BEST-HN(CO)CA and IPAP-(HA)CA(CO)NH sequences. Each spectrum was recorded for ~24 hours, so that recording for 48-72 hours would yield a further 1.4-1.7-fold increase in S/N. 2D H^N-C^α views of the separated 3D sub-spectra are shown in Figure 2.25. There are more peaks in the (HA)CA(CO)NH sub-spectra (~40) than in the BEST-HN(CO)CA sub-spectra (~30), demonstrating that while similar, the sensitivity of the (HA)CA(CO)NH sequence is slightly better. Some peaks are also much more intense in the (HA)CA(CO)NH

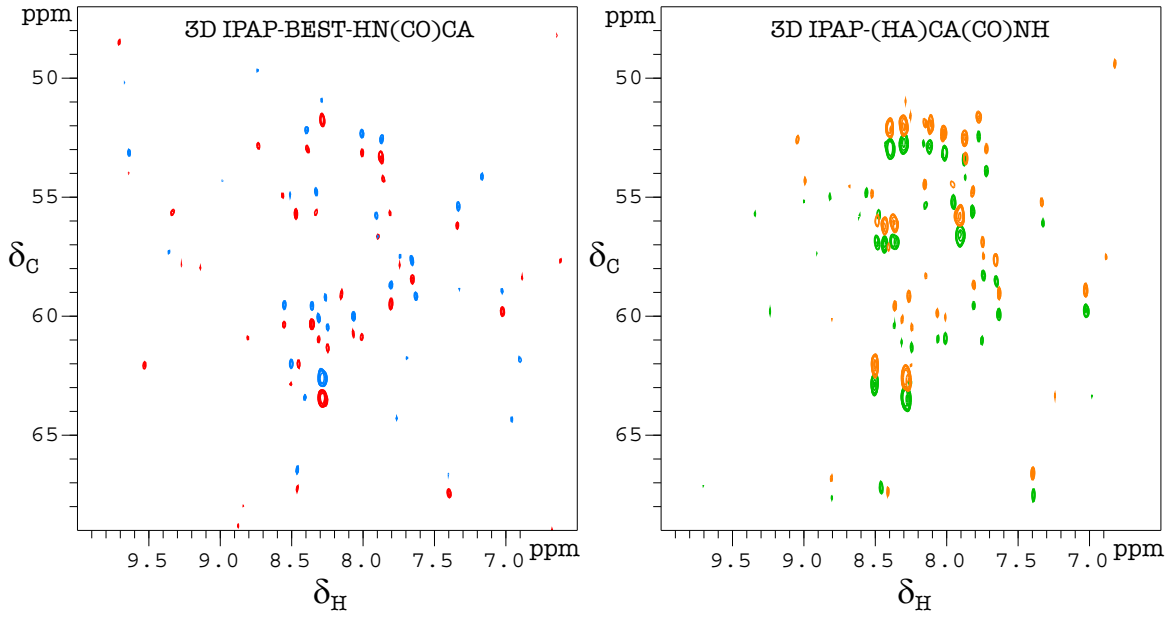

Figure 2.25: Comparison of 3D IPAP-BEST-HN(CO)CA (left panel) and 3D IPAP-(HA)CA(CO(NH) (right panel) spectra. Both spectra were recorded on 20 µM isolated L7/L12 (fraction 2) in phosphate buffer at 25 °C for 24 hours and contoured at the same level (5× RMS noise).

spectrum. These correspond to residues in the linker region, and are stronger in the (HA)CA(CO)NH spectrum because the magnetisation pathway starts on H^{α}. This means that the observed peak intensity is less affected by fast amide proton-water exchange, which will be significant for the unstructured, solvent-exposed linker residues. The number of detectable peaks in both spectra is still much lower than the total expected for L7/L12 (~120). Given that the spectra of L7/L12 on the ribosome are expected to be of poorer sensitivity than those of the isolated protein (due to the slightly faster relaxation of ribosome-bound L7/L12), particularly for aligned samples, the C^{α}-HSQC approach appeared most promising for measurement of C^{α}-H^{α} RDCs.

2.2.4.3 Assignment of C^{α}-H^{α} ^{13}C HSQC spectrum

One drawback of adopting the C^{α}-HSQC approach is that the H^{α}/C^{α} chemical shift assignments are required. These were not available for L7/12, as the previous NMR structure calculation relied on ^{15}N-labelled samples only [211]. Therefore, the H^{α} and C^{α} resonances were assigned using a concentrated (500 µM) sample of isolated L7/L12.

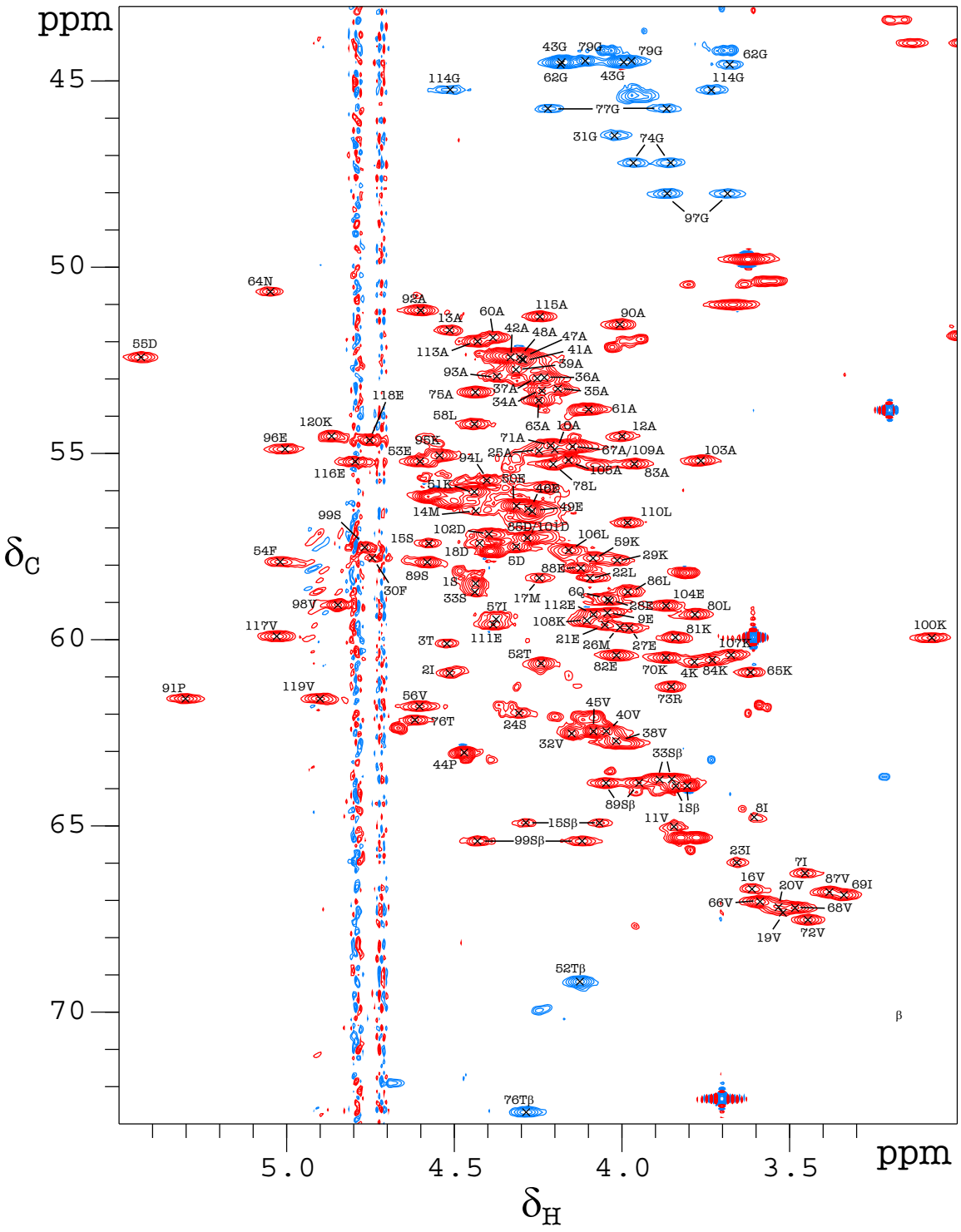

Figure 2.26: CT-^{13}C-HSQC spectrum showing the H^{α} and C^{α} chemical shift assignments of L7/L12. The glycine C^{α} and threonine C^{β} peaks (blue) appear with opposite sign to the remaining C^{α} peaks (red), because they have an even number of directly bonded aliphatic carbon atoms, all the other C^{α} and serine C^{β} nuclei have one directly bonded aliphatic carbon atom.

First, the assignments of the backbone N-H residues were confirmed and the C^{α}/C^{β} shifts measured using a standard set of triple-resonance experiments (HNCO, HNCA, HN(CO)CA, HNCACB, HN(CO)CACB & HN(CA)CO). The assigned ^{15}N HSQC spectrum is shown in Figure A.6. The assignments of residues 94 and 95 are swapped relative to the original assignment [211]. Second, the H^{α} shifts were obtained from HA(CACO)NH and HA(CA)NH experiments. The combined C^{α} and H^{α} shifts allowed complete assignment of the C^{α}/H^{α} region of the ^{13}C HSQC spectrum (Figure 2.26).

2.2.4.4 Comparison of C^{α}-H^{α} and N-H RDCs measured for isolated L7/L12 in phage

Both the isolated L7/L12 and the ribosome-bound L7/L12 were recorded in a deuterated buffer, which serves the purpose of reducing the line-broadening in the proton dimension that would result from unresolved H^{N}-H^{α} splittings. In addition, the spectrum recorded in D_2O-based buffer does not suffer from the presence of a noise stripe due to imperfectly suppressed water signal.

A lower concentration of phage was used for the measurement of C^{α}-H^{α} RDCs. This was to avoid severe J mismatch of the IPAP element, which arise due to a range of J+D splittings in combination with the restriction that only a single delay can be chosen for the IPAP element. The J mismatch can be greater for C^{α}-H^{α} couplings than N-H couplings because $D_{C^{\alpha}H^{\alpha}}$ is larger relative to $J_{C^{\alpha}H^{\alpha}}$ than for D_{NH}/J_{NH} for the same alignment strength. In other words, while $D_{C^{\alpha}H^{\alpha}} \sim 2D_{NH}$, $J_{C^{\alpha}H^{\alpha}}$ is only 1.5-fold larger than J_{NH}. In addition, because $D_{C^{\alpha}H^{\alpha}}$ is intrinsically larger than D_{NH}, relative weaker alignment strengths are required to obtain RDCs of sufficient size.

For both isotropic and aligned samples of isolated L7/L12, a 20-μM concentration was used to further assess the expected sensitivity for ribosome-bound L7/L12. The aligned sample was in the presence of 7.5 mg ml^{-1} phage, which gave rise to ~2.4 Hz splitting in the deuterium spectrum. IPAP C^{α}-HSQC experiments were recorded for a total of 35 hours and 80 hours for isotropic and aligned samples, respectively (aligned sub-spectra shown in Figure 2.27).

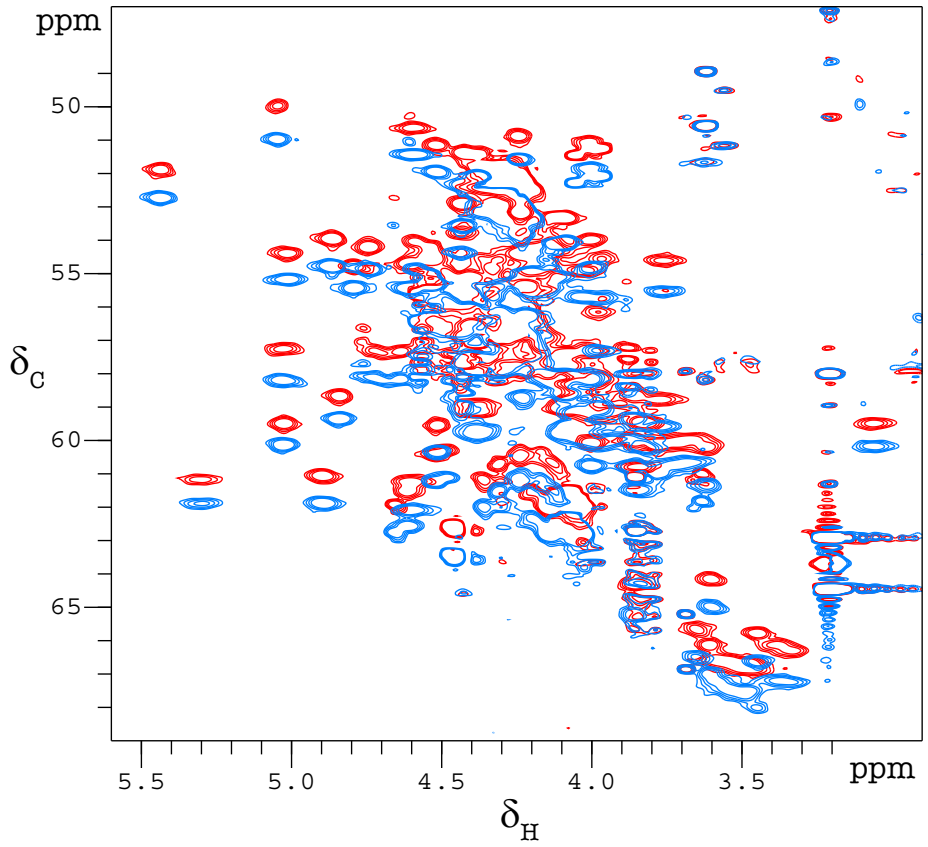

Figure 2.27: The aligned upfield (blue) and downfield (Red) C^{α}-HSQC sub-spectra, recorded on 20 µM isolated L7/L12 (fraction 2) in Tico buffer at 25 °C for 80 hours and contoured at 5× RMS noise.

Within both the isotropic and aligned IPAP C^{α}-HSQC spectra, 83 pairs of peaks out of 120 residues were assignable. After excluding residues with internal motion (as discussed in Section 2.2.3.4 for N-H RDCs), and overlapped peaks, 47 pairs of peaks were used for RDC analysis, of which 36 are from residues in the CTD, and 11 from residues in the NTD. Since the number of reliable RDCs for the NTD was low, together with the possible complications arising from oligomerisation processes, the NTD data was not analysed further.

Analysis of the data corresponding to residues in the CTD resulted a Q factor of 0.29, which indicates reasonable agreement between the data and the template structure (Figure 2.28 A), although it is not as good as for the N-H data (Q=0.19). The difference is partly due to the increased uncertainty in the measured splittings (RMS_{err}=1.26 Hz vs 0.18 Hz), but measurement uncertainty is still smaller than the RMSD between the measured and back-calculated RDCs (Table 2.7), indicating contributions from structural noise in the

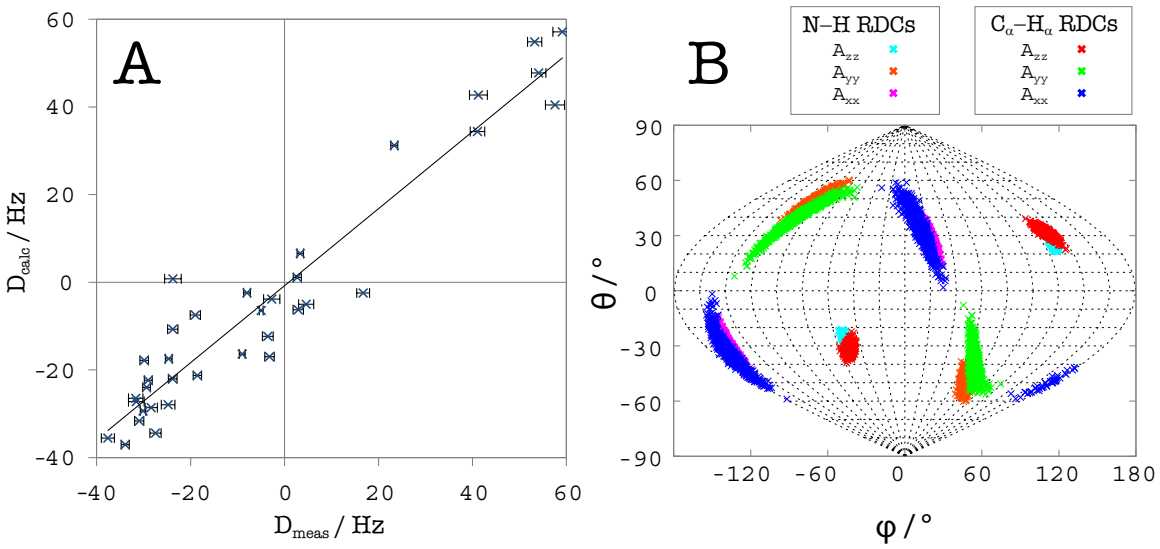

Figure 2.28: A: Correlation plot between the measured and back-calculated RDCs for the CTD of isolated L7/L12 (fraction 2). **B:** Sanson-Flamsteed projections showing the comparison of alignment tensor orientations derived for the CTD of isolated L7/L12 (fraction 2) from the N-H and C^α-H^α RDCs in phage.

template NMR structure.

The alignment tensor is very similar to that derived from the N-H data, as shown by overlaying the Sanson-Flamsteed projections (Figure 2.28 B). The calculated D_a^{NH}, is smaller than that for the N-H RDC data. This is a reflection of the decreased alignment strength, which is also reflected in the relative ^{2}H splittings (2.4 Hz vs 3.7 Hz). Both the rhombicity (R) and the orientation of the alignment tensor are similar.

N	RMS_{err}	RMSD	Q	D_a^{HN}	δD_a^{HN}	R	δR
36	1.26 Hz	8.6 Hz	0.29	-14.2 Hz	9.4 Hz	0.206	0.062

Table 2.7: Table of quality and alignment tensor parameters after SVD fitting of measured C^α-H^α RDCs to L7/L12 CTD structure for isolated L7/L12 in phage.

2.2.4.5 Approaches to measure C^α-H^α RDCs for ribosome-bound L7/L12

As for the isolated protein, C^α-H^α RDC data for ^{13}C,^{15}N labelled 70S ribosomes were acquired in deuterated Tico buffer. However, the C^α-HSQC spectrum was of much poorer quality than that of the isolated L7/L12 (Figure 2.30), with weaker peak intensities, loss of peaks, broader lines and increased overlap both among L7/L12 signals and with peaks

arising from the ribosomal RNA. As a result, only ~30 assignments could be transferred unambiguously from the isolated protein spectrum.

The corresponding C^{α}-IPAP-HSQC spectrum was measured for a total acquisition time of 42 hours. J couplings were calculated for 30 assigned peaks, with an average signal-to-noise ratio of ~15. The low signal-to-noise means there are large uncertainties in the measured couplings (RMS_{err}~5 Hz). This large uncertainty is reflected on the poor agreement with the J couplings measured for the isolated L7/L12, assuming that the residue-specific J couplings are relatively similar for both isolated and ribosome-bound L7/L12 (Figure 2.29).

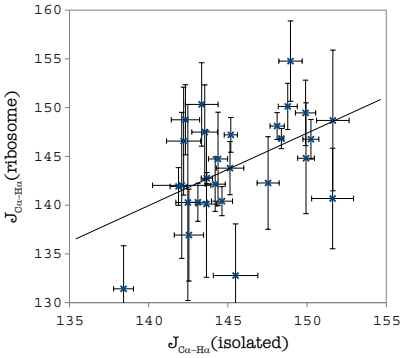

Figure 2.29: Correlation plot between the C^{α}-H^{α} J couplings of the CTD measured on isolated and ribosome-bound L7/L12.

2.2.4.6 Summary

The experiments discussed in this section demonstrate the challenge of measuring C^{α}-H^{α} RDCs on ribosome. The principal difficulties are sensitivity and peak-overlap within the ^{13}C HSQC spectrum. The data acquired on the ribosome under isotropic conditions were not of sufficient quality to proceed to measurement of the corresponding aligned spectrum.

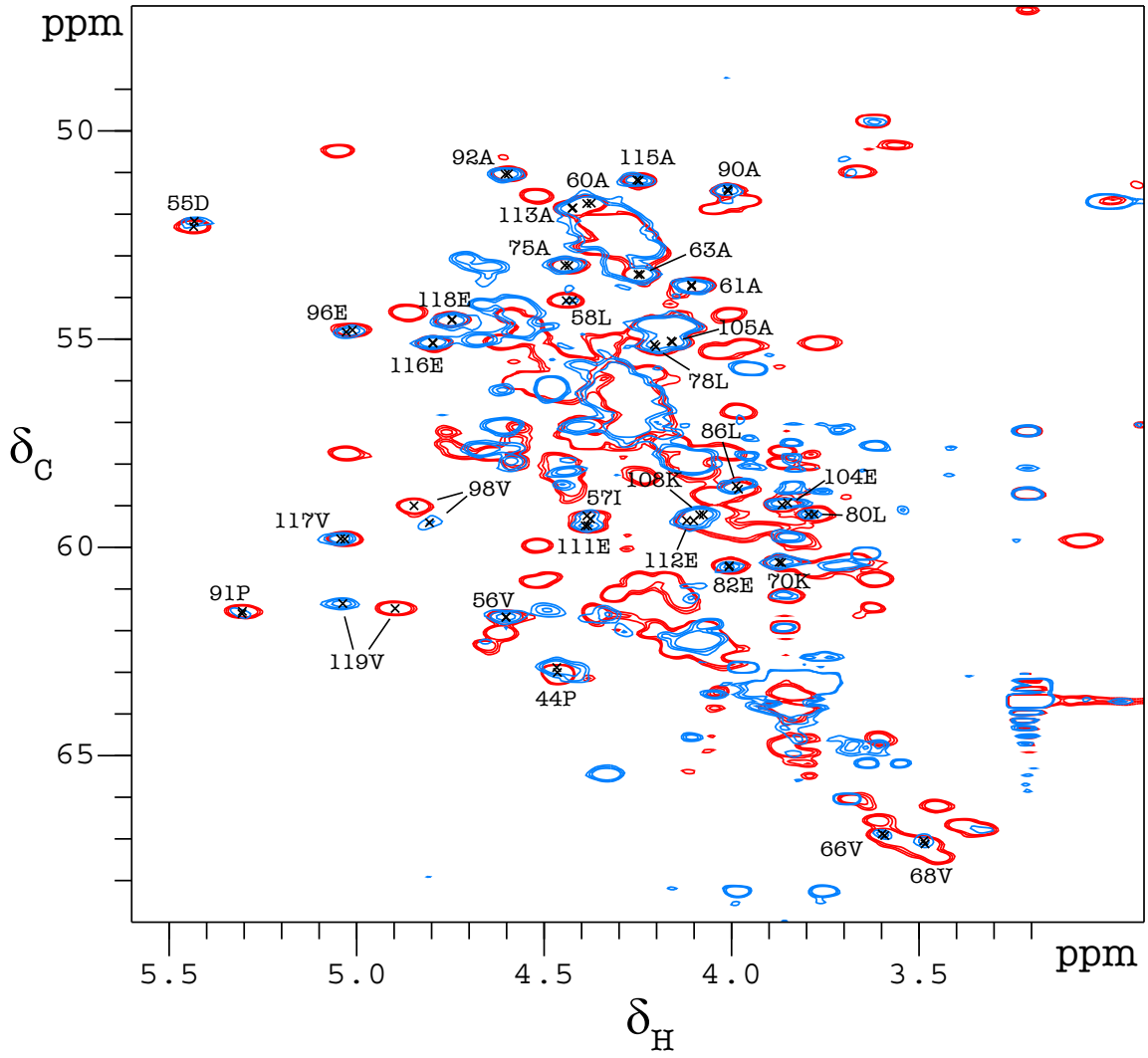

Figure 2.30: Comparison of C^{α}-HSQC spectra of ribosome-bound (blue) and isolated L7/L12 (red). The assignments that can be transferred unambiguously are labelled. The spectrum of ribosome-bound L7/L12 was recorded on 10 µM 70S ribosome and the spectrum of the isolated L7/L12 was recorded on 20 µM isolated L7/L12. Both spectra were recorded for 3 hours in Tico buffer at 25 °C and contoured at the same level (4× RMS noise).

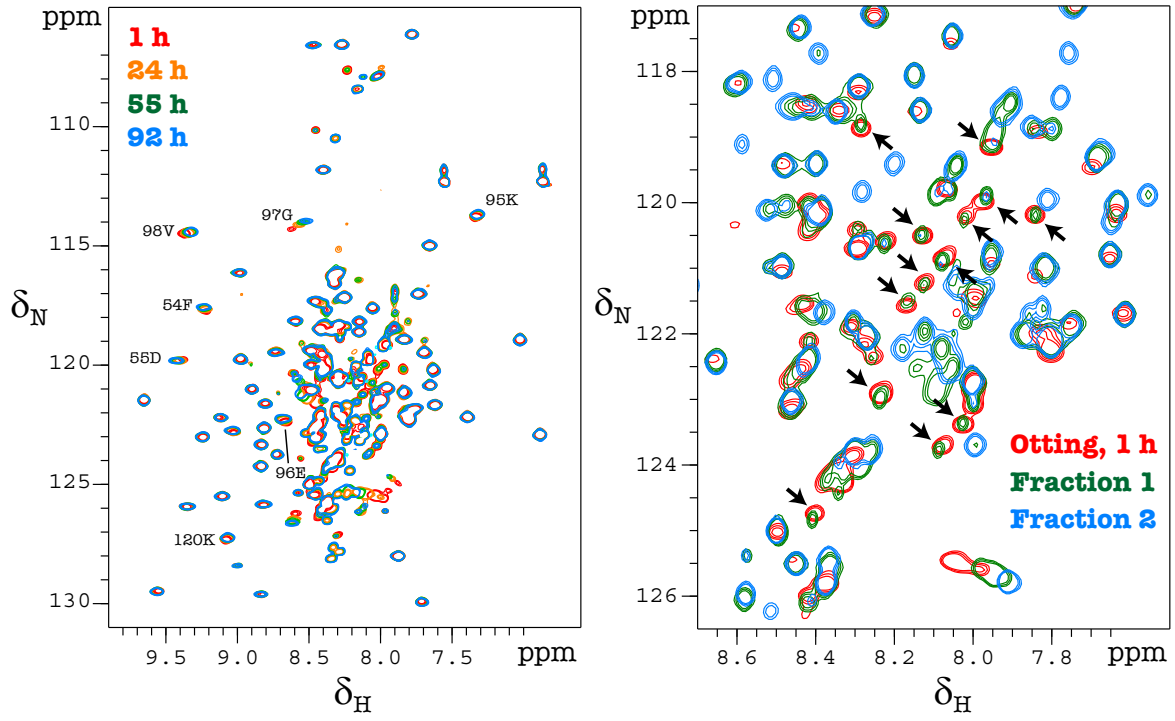

Figure 2.31: Left panel: Overlays of ^{15}N HSQC spectra of 250-µM isolated L7/L12 (fraction 2) in Otting medium recorded at 1 hour (red), 24 hours (orange), 55 hours (green) and 92 hours (blue) after sample preparation. CTD peaks showing chemical shift changes over time are labelled with their assignments. **Right panel:** Overlays of ^{15}N HSQC spectra of isolated L7/L12 (fraction 2) in Otting medium (red) and fraction 1 (green) and fraction 2 (blue) of isolated L7/L12 under isotropic conditions. The peaks of isolated L7/L12 (fraction 2) in Otting medium that overlay with the unassigned peaks in isotropic fraction 1 are highlighted with arrows.

2.2.5 Analysis of RDCs measured in Otting medium

As discussed in Section 2.1.3, measurement of RDCs in multiple alignment media is one way to reduce the degree of degeneracy in the relative orientations of bond vectors and the alignment tensor. The Otting medium (C12E5/*n*-hexanol) was tested for measurement of RDCs for isolated and ribosome-bound L7/L12. As described in the Introduction, Section 1.4.3, this liquid crystlline medium induces alignment on solely steric grounds, whereas the phage medium induces alignment mostly via electrostatic interactions. Using two media of different types is likely to yield different alignment tensors.

2.2.5.1 Approach to measure N-H RDCs for isolated L7/L12 in Otting medium

Initial experiments using the Otting medium were performed using the isolated L7/L12 protein. In principle, a single isotropic spectrum can be used as a reference spectrum together

with multiple aligned spectra in different alignment media. In other words, the isotropic splittings need only be measured once. The alignment of the sample was obtained by using 5 % $C_{12}E_5$/*n*-hexanol, resulting a deuterium splitting of 27 Hz (see Material and methods, section 5.2.2.2 for details of sample preparation). The N-H splittings was measured using the HSQC/TROSY approach.

In contrast to the phage-aligned spectrum, immediate loss of all NTD resonances was observed for the Otting-aligned spectrum. In addition, a slow change of shifts for a number of residues from the CTD was observed over a period of several days (Figure 2.31, left). Together these observations suggest that the N-terminal dimer may be partitioning into the bilayers formed by the polyethylene glycol molecules. The retention of peaks from the CTD indicates that the domain remains free, but the chemical shift changes suggest that there are residue-specific interactions with the alignment medium. The overall chemical shift perturbations are larger those observed in phage (Figure 2.32), leading to concern over possible structural perturbations induced by the Otting medium.

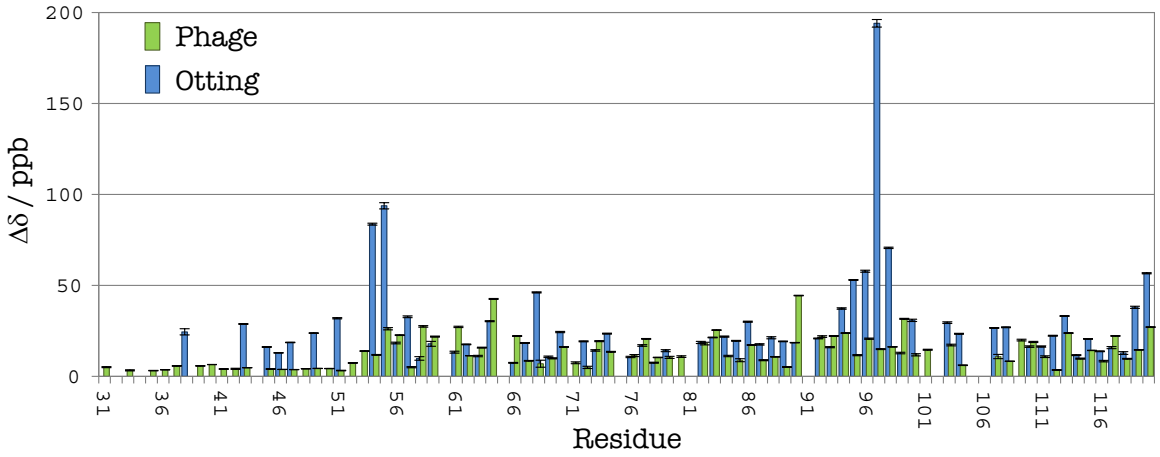

Figure 2.32: Chemical shift perturbations, $\Delta\delta$, of isolated L7/L12 upon addition of phage (green) and Otting medium (blue). Chemical shift changes in both the ^{1}H and ^{15}N dimensions were combined according to $\Delta\delta = \sqrt{(\Delta\delta_H)^2 + (\Delta\delta_N/5)^2}$.

After reaching equilibrium, as judged from the ^{15}N HSQC spectra, 44 splittings of the CTD residues were measured using the ^{15}N HSQC/TROSY approach. As expected, the orientation of the alignment tensor measured in Otting medium is very different to that in phage, as shown by the overlay of the Sanson-Flamsteed projections (Figure 2.33 B).

This difference can be quantified by calculating the angle between the two five-dimensional alignment tensors, which is given by the inverse cosine of their normalised scalar product. This gives an angle of 113 ° between the phage- and Otting-induced alignment tensors.

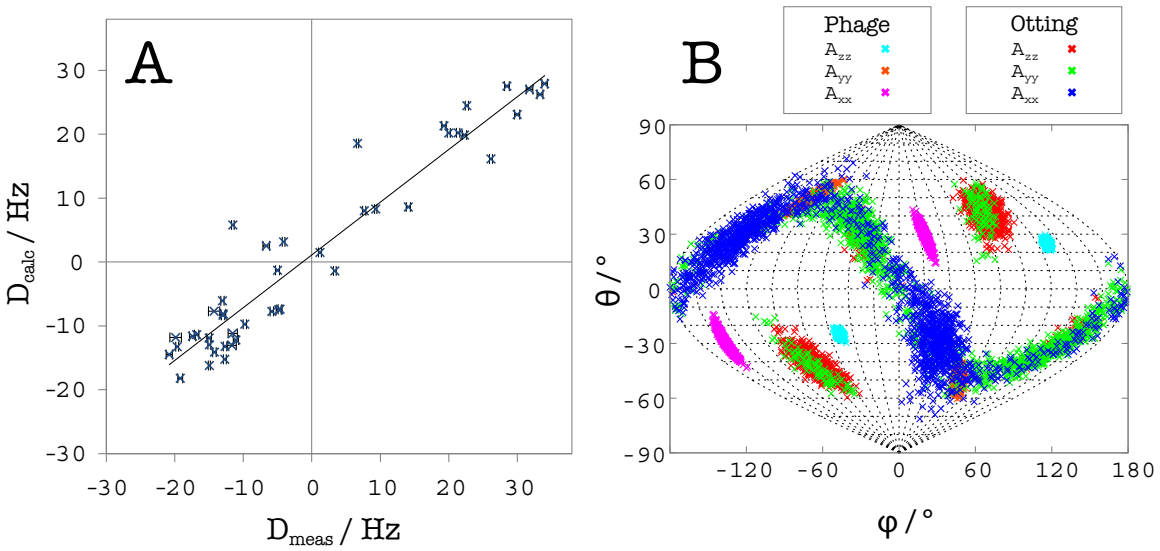

Figure 2.33: A: Correlation plot between the measured and back-calculated RDCs for the CTD of isolated L7/L12 (fraction 2) in Otting medium. **B:** Sanson-Flamsteed projection showing comparison of alignment tensor orientations derived from the CTD N-H RDCs of isolated L7/L12 (fraction 2) measured in phage and Otting media.

The Q factor and RMSD between the measured and back-calculated RDCs (Table 2.8) are poorer in Otting medium than in phage (Figure 2.33 A), with the RMSD twice that for the measurement in phage. However, there is only a small increase in the measurement uncertainty. This could be a reflection of possible structure perturbation to the CTD induced by interactions with the Otting medium.

	N	RMS_{err}	RMSD	Q	D_a^{NH}	δD_a^{NH}	R	δR
phage	50	0.17 Hz	2.9 Hz	0.19	-22.0 Hz	1.0 Hz	0.225	0.034
Otting	44	2.05 Hz	12.0 Hz	0.32	-14.0 Hz	9.4 Hz	0.275	0.503

Table 2.8: Comparison of quality and alignment tensor parameters after SVD fitting of measured N-H RDCs to L7/L12 CTD structure for isolated L7/L12 in phage and Otting media.

2.2.5.2 Approach to measure RDCs for ribosomes aligned in Otting medium

Alignment of ribosomes in Otting medium was also evaluated. In first instance, the same concentration of $C_{12}E_5$/*n*-hexanol medium (5 %) was used to align a 10-µM ribosome

sample. However, achieving alignment was more problematic than for the isolated protein, as reported by the measured deuterium splitting. Instead of an expected ^{2}H splitting of ~25-30 Hz, only 6.5 Hz was observed initially, with the peaks also showing broad linewidths. Temperature cycling up to 35 °C improved the splitting to 18 Hz, but the lines remained broad.

One pair of ^{15}N HSQC/TROSY spectra (~ 12 hours) was acquired prior to re-measurement of the ^{2}H splitting. An unusual pattern of peaks were observed in the ^{2}H spectrum, appearing as a doublet of doublets (Figure 2.34), which indicates possible phase separation of the medium. This situation was not resolved by repeated temperature cycling up to 35 °C.

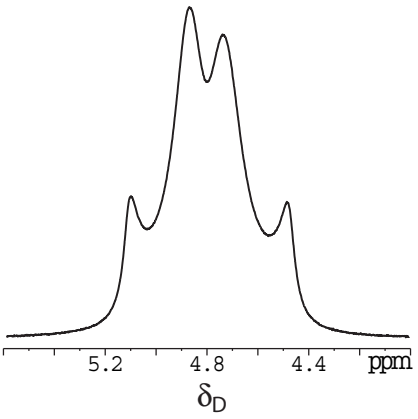

Figure 2.34: 1D ^{2}H spectrum of 10 μM ribosome sample in 5 % Otting medium after 19 h, showing a doublet of doublets.

It appeared that the ribosome was interfering with the formation and alignment of the liquid crystal phase. As described in [134], the spacing between the stacked bilayers is estimated to be only a few nm at 5 % (w/w) concentration. It is possible that there is simply not enough space for the ribosomes in the inter-bilayer space (ribosome diameter ~25 nm). This hypothesis was evaluated by diluting the sample to reduce the concentration of the Otting medium to 3 %. The ribosome concentration is also diluted to ~6 μM. After the dilution of the sample, the deuterium spectrum showed a splitting of 15 Hz and had narrower linweidths (although not as narrow as for the sample of isolated L7/L12), which remained stable over time.

Following the ~24 hours of data acquisition described above, the measured diffusion of the sample indicated significant degradation of the ribosome (to a degree usually observed after three days of measurement). The faster degradation of the ribosome sample was probably due to the extensive sample manipulation and temperature cycling performed during the process attempting to obtain stable ^{2}H splittings. Unfortunately, the degradation of the sample meant that it was not possible to continue with the measurement of the N-H

splittings.

2.2.5.3 Summary

Alignment of isolated and ribosome-bound L7/L12 with Otting medium was not as successful as alignment with bacteriophage. Isolated L7/L12 appears to show strong interactions with the Otting medium, with the N-terminal dimer domains suspected to completely partition into the polyethylene glycol bilayers, and the CTD showing several residue-specific interactions. The ribosome appeared to interfere with formation of a homogeneous well-aligned liquid crystal phase, but this could be because the concentration of the Otting medium was too high at 3-5 % (w/w). The typical phage concentration used for ribosome alignment was ~10 mg ml^{-1} (or 1 %), so alignment of the ribosome with the Otting medium may be successful at lower concentration.

2.2.6 Approaches to measure other types of RDCs

As discussed in the previous section, the difficulty of measuring NH RDCs in Otting medium emphasises the requirement of measuring multiple types of RDC in phage. These RDCs can be measured for other couplings in the peptide plane.

2.2.6.1 Approaches for measurement of H-C' and N-C' RDCs on the ribosome

Measurement of H-C' and N-C' RDCs has the advantage that these two RDCs can be measured simultaneously in a 2D N-H correlation spectrum with relatively high sensitivity. However, the sizes of these two couplings are relatively small compared to N-H RDCs, being ~3-fold and ~8-fold smaller, respectively.

The simplest method to measure these two couplings is to remove the ^{13}C' decoupling pulse from the indirect evolution time in a ^{15}N HSQC or TROSY sequence. As well as generating a splitting in the indirect ^{15}N dimension, the resulting doublet components are also displaced in the acquisition proton dimension. This is due to the natural E.COSY effect, which arises because there is no change in the C' spin-state during the back-transfer of

magnetisation from ^{15}N to ^{1}H. The N-C' and H-C' splittings are then measured from the displacement of the doublet components in the ^{15}N and proton dimensions, respectively. However, this requires spectra of very high resolution to separate the E.COSY components. This problem can be avoided by recording the data in an IPAP fashion. The disadvantage of the IPAP approach is that the long constant-time period required for the IPAP element ($1/2J_{\mathrm{N\text{-}C'}}$~30 ms) causes sensitivity losses.

To evaluate if the precision of measurement will be sufficient relative to predicted range of RDC values to allow meaningful interpretation of the data, IPAP-E.COSY(C')-^{15}N-HSQC-TROSY spectra were measured on a 10-μM ^{13}C, ^{15}N labelled ribosome sample. Splittings were extracted from combined spectra representing a total acquisition time of 40 hours (spectra are shown in Figure A.7 on page 217).

Peak positions, signal-to-noise ratios and linewidths were analysed for 75 peaks. The calculated RMS uncertainties are 1.9 Hz and 0.9 Hz for $J_{\mathrm{H\text{-}C'}}$ and $J_{\mathrm{N\text{-}C'}}$, respectively. These uncertainties should be compared to the predicted range of RDCs. Assuming $D_{\mathrm{a}}^{\mathrm{NH}}$ would be the same as previously obtained for ribosomes aligned in phage (20 Hz), the magnitudes of the expected couplings would fall in the range 0-12 Hz and 0-5 Hz for H-C' and N-C', respectively. Assuming that the aligned splittings can be measured with the same precision as the isotropic splittings would yield estimated uncertainties of 2.7 Hz and 1.2 Hz for $D_{\mathrm{H\text{-}C'}}$ and $D_{\mathrm{N\text{-}C'}}$, respectively. Therefore, the predicted uncertainties are ~20 % of the maximum expected coupling for both the H-C' (2.7/12=0.22) and N-C' (1.2/5=0.24) coupling types. This level of fractional uncertainty corresponds to a Q factor of 0.45-0.5, indicating that while in isolation the utility of these couplings would be limited, they could still be useful when combined with the N-H couplings.

However, in practice, the proton linewidth for the aligned sample will be significantly broader than for the isotropic sample due to the proton-proton dipolar couplings, and hence the uncertainties in the H-C' RDCs may be too large for this coupling to be useful.

In conclusion, it may be possible to measure the N-C' couplings at sufficient precision to provide useful information, but the H-C' couplings will be less precise due to broad

linewidths in the aligned sample and are unlikely to be useful.

2.2.6.2 Approach to measure C'-C^{α} RDCs

The final type of coupling considered in this work was the C'-C^{α} couplings. The magnitude of this coupling is one-fifth of the N-H coupling, which is in between the magnitude of the H-C' and N-C' couplings. The measurement of the C'-C^{α} coupling via the ^{13}C HSQC spectrum is associated with the same difficulties as described for H^{α}-C^{α} couplings, including peak overlap and poor resolution due to the limited ^{13}C acquisition time. The latter problem is even more important for measuring C'-C^{α} couplings, since the size of the RDCs is 10 times smaller than those for the H^{α}-C^{α} coupling (hence requiring narrower linewidths to allow measurement with sufficient precision).

An alternative approach is to measure the splittings in a 3D fashion, using an HNCO-type sequence. The HNCO-type experiment is approximately three times more sensitive than the HN(CO)CA sequence tested during the attempt to measure H^{α}-C^{α} splittings, and hence might be feasible for use with ribosome samples. The advantage is that a longer acquisition time can be used on C' than C^{α}, since the C' acquisition time is not limited by evolution of other carbon-carbon couplings.

To evaluate the sensitivity and resulting uncertainties of this approach, a 3D in-phase spectrum using an IPAP-BEST-HNCO sequence was measured for 12 hours on a 20-μM isolated L7/L12 sample. The 2D H-C' projection of the 3D in-phase HNCO spectrum is shown in Figure 2.35, left panel. Peaks representing 80 % of the residues were picked for the signal-to-noise ratio and linewidth analysis. The average signal-to-noise ratio is 10 with an average linewidth of 27 Hz. Extending the total acquisition time to 72 hours would improve the signal-to-noise ratio to 25, which would then yield an estimated peak position uncertainty of 0.5×(27/25) = 0.5 Hz. Assuming a similar signal-to-noise ratio could be obtained for the ribosome sample leads to a predicted uncertainty in the measured RDCs of ~1 Hz. Again assuming an alignment strength corresponding to a D_a^{NH} of 20 Hz, the predicted maximum RDC for the C'-C^{α} coupling would be 8 Hz. Therefore, the estimated

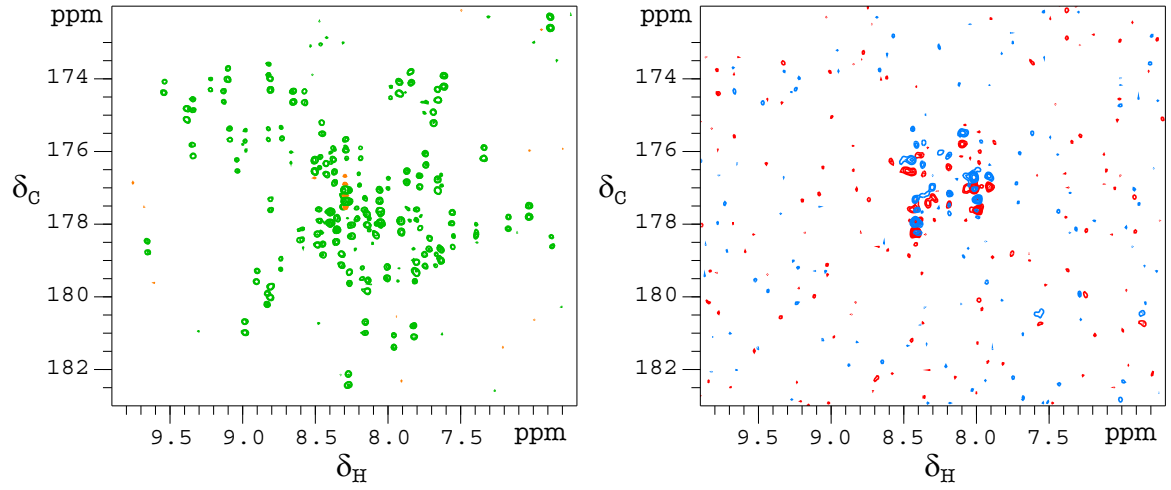

Figure 2.35: Left panel: 2D H-C' projection of the 3D in-phase HNCO spectrum of isolated L7/L12. Spectrum recorded on 20 µM isolated L7/L12 (fraction 2) and contoured with 5× RMS noise. **Right panel:** 2D H-C' IPAP-BEST-HNCO spectrum of ribosome-bound L7/L12 (10 µM 70S ribosomes) with 3× RMS noise. Both spectra were recorded for 12 hours in Tico buffer at 25 °C.

uncertainty is approximately 12 % of the maximum expected RDC, which corresponds to Q factor of 0.25-0.3, which is sufficiently small to be useful.

On the basis of the above analysis, an attempt was made to measure the C'-C^{α} splittings for a phage-aligned ribosome sample using the IPAP-BEST-HNCO sequence. The experiment was acquired as a 2D H-C' spectrum to maximise the sensitivity. The measured spectrum for 12 hours total acquisition time is shown in Figure 2.35, right panel. Only a few peaks were observed, which correspond to peaks from the linker region and other intense unassigned peaks in the ^{15}N HSQC spectrum of the ribosome.

In conclusion, this approach appears not to be feasible due to the low sensitivity of the spectrum measured on the ribosome. The difference in sensitivity between the isotropic spectrum of the isolated L7/L12 and the phage-aligned spectrum of ribosome-bound L7/L12 is the result of both the increased relaxation rates of ribosome-bound L7/L12 and the decrease in the signal-to-noise ratio due to the increased proton linewidths in the aligned sample (unresolved proton-proton couplings). The triple-resonance experiments are more sensitive to small differences in relaxation rates than simple 2D experiments due to the long delays and multiple transfer steps in the pulse sequence of the triple-resonance experiments.

2.3 Conclusions

The isolated L7/L12 protein shows complicated oligomerisation behaviour in solution, reflected in the multiple fractions observed under gel filtration. Further analytical gel filtration and NMR diffusion and ^{15}N HSQC experiments suggested that fraction 1 is higher order oligomeric form of the L7/L12 protein, which does not exchange easily with fraction 2. Within fraction 2, monomeric, dimeric and possible tetrameric forms all appear to exist in concentration-dependent equilibria, and the different oligomeric states are in fast exchange on the NMR timescale.

N-H RDCs were successfully measured on both isolated and ribosome-bound L7/L12 aligned with phage. The alignment tensors of the CTDs of the two fractions of the isolated protein were very similar. This is at least partly because the alignments of the NTDs are very weak, and thus any influence on the alignment of the CTD from the NTD alignment will be small. In addition, within the phage medium, the alignments of each the domains arise mainly from the interaction between the domains themselves and the alignment medium, such that any relative orientation preference existing between the domains under isotropic conditions is likely to be swamped. In slight contrast, the alignment tensors of the isolated and ribosome-bound L7/L12 CTDs were found to be more different than those of the fractions of isolated L7/L12. Although the difference was not highly significant compared to the estimated uncertainties, the result might hint at propagation of partial alignment through the linker region, such that the (unknown) alignment of the ribosome core particle is influencing the anisotropic orientational distribution of the ribosome-bound L7/L12 CTD. The structural refinement of the CTD in its ribosome-bound form from the RDC data revealed a very similar structure to that of the isolated protein, confirming that the C-terminal domain structure remains the same both on and off the ribosome.

It has been proposed that the linker region of L7/L12 can exist in both an extended and a compact, part-helical form, such that either one or the other CTD (but not both) can be retracted close to the NTD at any one time. This behaviour may explain why only two copies of the CTD appear to contribute to the L7/L12 NMR signal of the ribosome, which

contains four copies of the L7/L12 protein overall. This picture then implies that the NMR-observable CTDs belong to different L7/L12 dimers on the ribosome, and hence must be in different positions on the L10 scaffold protein. As such, they should give rise to two sets of RDCs, but only a single set was detected. Therefore, either the differences were too small to be detected, or the L7/L12 dimers exchange rapidly between the two binding positions on L10.

Attempts to measure other types of RDC were made to allow more powerful structure refinement of both isolated and ribosome-bound L7/L12, and to continue the development of the RDC approach for subsequent application to RNCs. For the measurement of C^{α}-H^{α} RDCs, existing pulse sequences were tested and modified to optimise sensitivity and resolution for application to the ribosome-bound L7/L12. Using an IPAP-C^{α}-HSQC sequence for the measurement of C^{α}-H^{α} RDCs was successful for the isolated protein, but measuring C^{α}-H^{α} RDCs on the ribosome was found to be more challenging, with the main difficulties being low sensitivity and peak overlap in the ^{13}C HSQC spectrum. Other types of RDCs were also considered, including H-C', N-C' and C'-C^{α} RDCs, and while it was shown that it may be feasible to measure the N-C' RDCs on the ribosome with sufficient precision to provide useful information, meaningful measurement of the H-C' and C'-C^{α} couplings would be infeasible due to broad proton linewidths in the aligned sample for the H-C' coupling and low sensitivity for the C'-C^{α} coupling.

In addition, a different alignment medium, the Otthing medium, was also investigated for the alignment of isolated and ribosome-bound L7/L12. This approach was not as successful as alignment in phage. The isolated L7/L12 protein showed strong interaction with the Otting medium, principally via the N-terminal dimer domain. The alignment of ribosomes in Otting medium was hampered by interference of the ribosome with the formation of the liquid crystal phase, although improvements are potentially achievable by reducing the concentration of the alignment medium.

In summary, the work presented in this chapter has demonstrated the feasibility of measuring RDCs on ribosome-bound proteins, and has provided the first direct structural

information on the L7/L12 protein in its ribosome-bound state.

Chapter 3

Application of RDCs for characterisation of ribosome-nascent chain complexes (RNCs)

As described in Chapter 2, solution-state NMR, and in particular RDCs, can be used to observe and characterise the dynamic stalk region of the ribosome. Recent studies have shown that NMR can be applied to RNCs of several proteins [125, 126, 225]. This is possible because, despite being tethered to the large ribosome particle, the nascent chain is inherently dynamic.

Present NMR strategies for evaluating the structural and dynamic properties of ribosome-bound nascent chains (RNCs) include evaluating HSQC spectra of the corresponding, released proteins and assigning resonances in RNC spectra by comparative inspection. Chemical shift differences between the ribosome-bound NC and the corresponding isolated protein can then be used to provide information about the extent of folding. For resonances arising from regions of disorder, N-H chemical shifts act as highly sensitive probes, while for dispersed resonances, i.e. those corresponding to persistent tertiary structure, the use of selectively protonated methyl groups within an otherwise deuterated background significantly improves the sensitivity relative to simple ^{13}C-labelled, protonated sample [226]. Chemical shift analysis is an excellent probe for the formation of structure, but it

is typically unable to provide direct structural information. Traditionally, NOE experiments are used to provide distance-restraints within proteins [227], but the achievable ribosome concentration of 10 μM combined with limited sample lifetimes means that such experiments are infeasible. In order to begin to define the orientation and structural preference of RNCs during co-translational folding, this chapter discusses the application of the RDC approach that was developed for the 70S ribosome in Chapter 2.

3.1 Introduction

3.1.1 Generation of ribosome-nascent chain complexes for structural studies

Newly synthesised polypeptide chains are released by the ribosome at the termination of an mRNA transcript, which is usually signalled by the presence of a stop codon. While clearly the release of proteins into the cellular milieu is necessary for them to carry out their biological function, it introduces a significant challenge for the isolation of RNCs in studies of protein folding on the ribosome.

Several methods have been developed to produce RNCs with homogeneously stalled nascent chains [97] for both biochemical and biophysical studies. The most well established means of stalling nascent chains (NCs) on their ribosomes has been obtained by using *in vitro* transcription-translation (cell-free) systems, in which stalling is achieved by terminating translation using linearised mRNA transcripts in the absence of a stop codon [225, 228]. Within the cell-free system all the components can be controlled and the stalling of the NC can be achieved relatively simply. An additional consideration arises due to the the L7/L12 resonances, which are likely to dominate the spectra as there are two copies of NMR-visible L7/L12 compared to a single nascent chain per ribosome. Therefore, high NC labelling efficiency is essential to enable the maximum NMR signal to be obtained. Cell-free systems can be used for generating selectively-isotopically labelled RNCs that are suitable for NMR by supplementing the system with isotopically labelled amino acids [225]. This method is

efficient, although it is costly to produce the relative large quantities of RNCs needed for NMR experiments.

As an alternative to cell-free systems, recent advances in *in vivo* methods have enabled the production of RNCs [124, 125, 229, 230]. These methods allow the RNCs to be generated relatively inexpensively. However challenges arise from the inherent complexity of the strategy, which requires tight control of the translational process as it occurs within the cell. In this case, specific peptidic sequences are engineered at the C-terminus of the DNA sequence to induce translational arrest. In particular, the use of the secretion monitor protein, SecM [231] and also more recently, the tryptophanase operon-derived sequence, TnaC peptide [25] has made *in vivo* RNC production possible [124–126, 230].

SecM is a 170-amino acid *E.coli* protein that regulates the expression of a downstream protein, SecA ATPase within the secM-SecA operon [232, 233]. A 17-amino acid motif from the C-terminus of SecM was found to induce elongation arrest, nine amino acids of which are required for the stalling mechanism (shown in bold): 150-**F**XXXX**WI**XXXX**GIRAGP**-166 [231]. It has been suggested that the stalling motif is recognised by regions of ribosomal proteins L4 and L22, and also by parts of the 23S ribosomal RNA (Figure 3.1) [51, 231, 234], and that this results in stalling of translation at Gly-165, producing a peptidyl-glycyl-tRNA-nascent chain complex which is trapped at the P-site of the peptidyl transferase centre of the ribosome. It has also been noted that the following codon, Pro-tRNA^{166}, also appears to be essential for arrest through its effects at the A-site [51].

Cryo-EM studies have revealed a possible mechanism for SecM stalling, in which interaction of the SecM peptide in the exit tunnel with the 23S rRNA induces a conformational change in the 23S rRNA. This is propagated to the P-site 3'-amino acylated A nuclotide, pushing the carbonyl-carbon of the attached Gly-165 residue away from the amine group of the Pro-tRNA^{166} (Figure 3.1), preventing peptide bond formation and subsequent translocation.

The *in vivo* approach whose development and application is described in this chapter enables the RNC to be produced within the natural cellular environment in the presence of

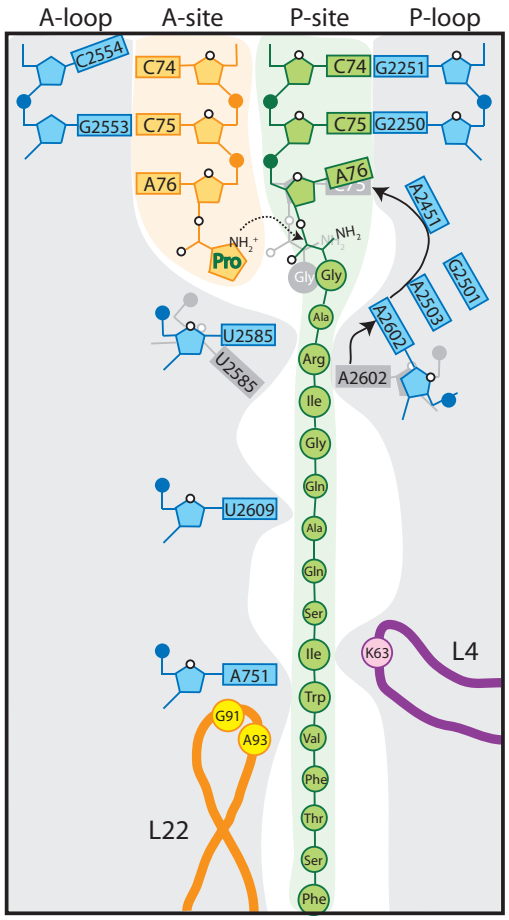

Figure 3.1: Schematic of the conformational changes of the ribosome rRNA and the P-site tRNA induced by the SecM sequence. Where SecM nascent chain is in green, ribosomal proteins L4 in purple, L22 in orange. Produced according to [51].

auxiliary factors such as chaperones. The advantage of the method for producing SecM-stalled RNCs is that it enables production of pure and homogeneous ribosome-nascent chain complexes with a minimum degree of ribosome background labelling.

Another important aspect of the study of RNCs by NMR concerns monitoring their integrity, as they are dynamic complexes of finite lifetime. This is imperative to ensure that the signal observed arises from a ribosome-bound NC and not from released species. The released NC is much smaller and more flexible than the intact complex and is therefore less motionally restricted. Since NMR sensitivity is related to the tumbling rate of a molecule, released NCs will contribute more to the observed signal than attached NCs.

3.1.2 NMR investigations of ddFLN$_{646\text{-}838}$ RNC (Dom5+110)

To study protein folding on the ribosome, the protein ddFLN was selected. ddFLN, also known as ABP-120, is derived from *Dictyostelium discoideum*. It is an actin binding protein, which consists of one N-terminal actin-binding domain and six repeats of immunoglobulin domains (1-6). For these studies, domains 5 (Dom5) and 6 (Dom6) of ddFLN were selected, which are flanked by the domain boundaries at residues 646 and 850, using standard sequence numbering [235]. In the construct used in this study, the terminal beta strand in domain 6 was removed, which means that it is unable to acquire its native fold and remains unstructured. As an isolated protein, ddFLN has been well characterised with X-ray crystal structures available for domains 4-6 [236] and domains 5-6 [235], as well as NMR assignments of both the native and urea-denatured domain 5 in solution [237]. ddFLN has also been used in previous NMR studies of co-translational folding of RNCs, which demonstrated that domain 5 can acquire its native fold as an attached nascent chain with truncated domain 6 acting as a tether to the ribosome (Figure 3.2 B) [124, 225]. Recently, an approach for generating RNCs of ddFLN in *E. coli* has been established. In this study, a construct of ddFLN was designed in which Dom5 and Dom6 were stalled using a SecM sequence at the C-terminus of the protein (Figure 3.2). The RNC of ddFLN is termed here as Dom5+110, which reflects the distance in residues at the end of Dom5 and the PTC, in this case 110 amino acids.

At present, there is very little three-dimensional structural information on ribosome-bound nascent chains. RDC measurements, as shown in Chapter 2, can provide details on the structural preferences of the dynamic L7/L12 region on the ribosome and offer the prospect of deriving structural models. This chapter describes a study in which an RDC approach applied to RNCs. This chapter explores both the approaches to produce RNCs and characterise their biochemical and NMR properties as well as the application of the RDC methodology described in chapter 2.

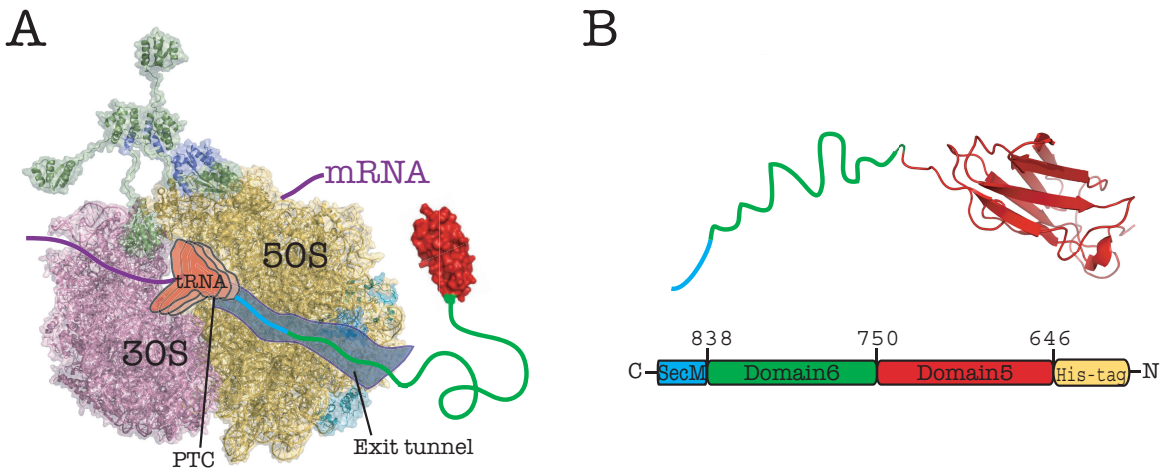

Figure 3.2: A: Schematic of the Dom5+110 RNC. The crystal structure of the 70S ribosome, with the PTC highlighted by the presence of tRNAs. Emerging from the exit tunnel is a nascent chain of Dom5+110; the SecM-Dom6 linker is modelled in blue/green, while Dom5 is shown in red (based upon PDB ID 1QFH). **B:** Cartoon representation of Dom5+110 construct, showing the folded Dom5 (red) and truncated Dom6 being unstructured (green).

3.2 Results and Discussion

3.2.1 Optimisation of the production of *in vivo*-derived RNCs suitable for RDC studies

Central to the study of RNCs using RDCs is the need for suitable samples. To further develop existing *in vivo* production strategies, the expression and purification of RNCs was investigated. A detailed description of the RNC production strategy is given in section 5.1 on page 182 and the most recent developments are summarised in the following section.

3.2.1.1 Development of an expression strategy for *in vivo* RNCs

The original RNC production strategy is shown in Figure 3.3 A. To achieve the high concentration of ribosomes required for expression of RNCs, *E. coli* BL21 (DE3) are grown to a very high cell density (typically OD_{600}=4 or higher) in MDG medium (based on an auto-induction medium [238]) for 16-20 h. It was found that growing the cells at 30 °C instead of 37 °C typically yielded a higher cell density, up to OD_{600}=6 compared to ~2. The end point OD_{600} of the cells grown in MDG medium was typically used as an indication of cell viability, where cultures with OD_{600}<2 were discarded.

After growth, the cells were harvested and transferred into expression medium that did not contain any carbon or nitrogen sources (Figure 3.3 A). The cells were incubated in this medium for a short period of time (~25 mins) to allow depletion of all unlabelled carbon/nitrogen sources carried over from the MDG medium. At the end of the depletion period (indicated by a decrease in the OD_{600} reading, Figure 3.3 A), isotopically labelled carbon/nitrogen sources were added together with IPTG for expression.

Originally, minimal medium (M9) was used, which gave a yield of 100 pmol per litre. Repeated expression attempts did not result in an overall improvement in the yield, and therefore enhanced minimal media (EM9) was introduced as the expression medium. The level of RNCs expression was improved significantly, by an order of magnitude to >500 pmol per litre. At these expression levels, a band corresponding to the NC was easily observed by silver-stained SDS-PAGE. One minor problem was that the cells did not reach complete depletion as indicated by a drop in OD_{600} for more than 1 hour, which is probably due to the rich components of the medium. The presence of any residual ^{14}N or ^{12}C during expression can lead to a decrease in NMR observable signals from the RNC. Instead, to rule out that the improved expression levels did not arise from a carry-over of nitrogen/carbon sources, a washing step (Figure 3.3 B) was introduced in place of the depletion stage to remove the residual unlabelled nitrogen/carbon sources from the expression medium.

In addition to the washing step, which enables 100 % labelling of the NC, rifampicin was added as a supplement to the EM9 medium to prevent background labelling of the ribosomes. Rifampicin is a bactericidal antibiotic that inhibits the natural bacterial RNA synthesis by inhibiting bacterial DNA-dependent RNA polymerase [239]. It binds to the RNA polymerase at a site next to the active centre and physically arrests RNA production [240, 241]. As the expression of the NC is under control of the T7 polymerase, it is not affected by rifampicin; adding rifampicin (150 mg l^{-1}) to the expression medium after induction was found to reduce the expression — and therefore labelling — of *E. coli* proteins, including ribosomal proteins, by *ca.* 50 % (Figure 3.4).

Although the use of EM9 medium improved the expression level of RNCs, amino-acid

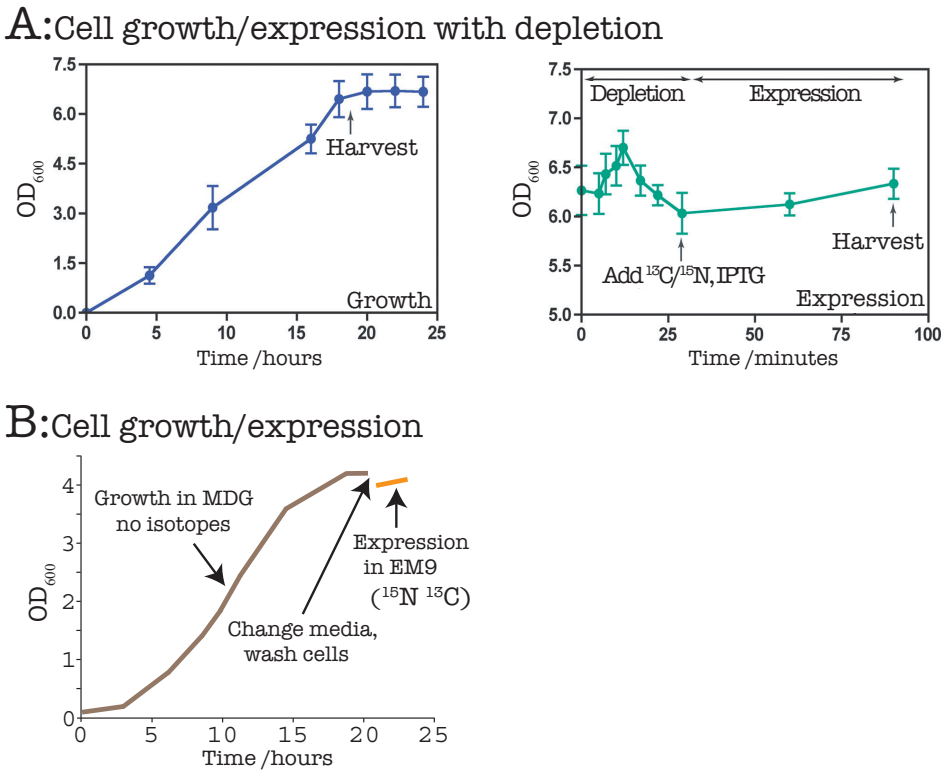

Figure 3.3: **A:** The original growth and expression of RNCs using the depletion method at the expression stage. **B:** The optimised protocol for the growth and expression of RNCs to achieve unlabelled ribosomes and labelled NCs.

supplements (Cambridge Isotope Laboratories, Inc.) were also evaluated in the effort to further improve RNC expression levels. When added to EM9 medium, they had previously been found to improve the yield of ddFLN-RNCs, enabling short NC-length constructs (< 150 residues) to express at levels comparable to the full-length (220 amino acids) NC (Marilia Karyadi, personal communication). In the current study, the amino-acid supplements were used during the expression of wild-type (WT) α-synuclein (αSyn) RNCs, whose yield is typically five times lower than that of the ddFLN RNCs produced under the same expression conditions. The amino-acid supplements were found to improve the yield of the WT-αSyn RNC by three-fold (from 480 to 1500 pmol per litre, Annika Weise, personal communication).

The improvements that were developed during this study increased expression levels by ~30 fold, and enabled RNCs suitable for RDC experiments to be produced.

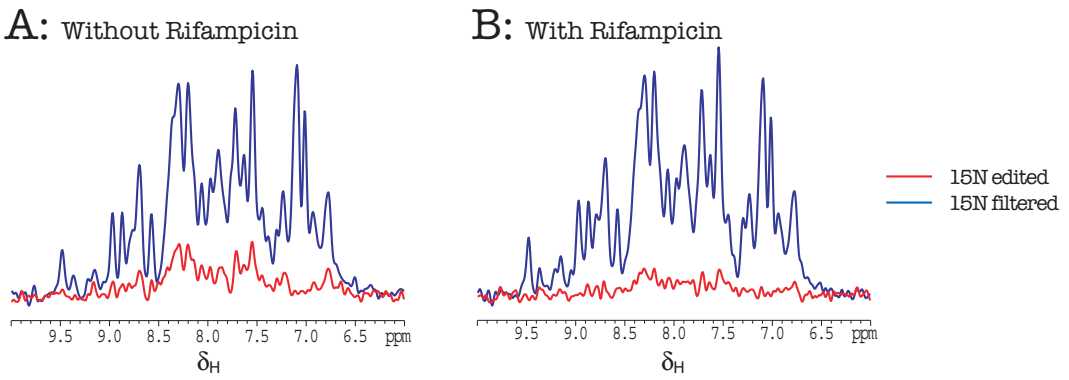

Figure 3.4: Comparison of the amount of background labelling arising of ribosome in the absence (A) and presence (B) of rifampicin produced under RNC expression conditions. The ^{15}N-edited spectra (red) contain signals solely from ^{1}H attached to ^{15}N and the ^{15}N-filtered spectra contain signals from ^{1}H not attached to ^{15}N. **A:** In the absence of rifampicin, the intensity of the ^{15}N-edited spectrum is 24 % of the ^{15}N-filtered spectrum, indicating 20 % labelling of the 70S N-H groups with ^{15}N. **B:** In the presence of rifampicin, the intensity of the ^{15}N-edited spectrum is 11 % of the ^{15}N-filtered spectrum, which means that 10 % of the 70S ribosome N-H groups are labelled with ^{15}N.

3.2.1.2 Development of the purification strategy for *in vivo*-derived RNCs

A critical step in RNC production is the separation of NC-occupied ribosomes from the released NCs, empty ribosomes and other endogenous proteins. During this work, the purification scheme described in [124] was extensively modified and improved.

There are two key requirements for RDC measurements of RNCs: that the samples are homogeneous, as impurities result in reduced sample lifetimes, and also that they are of high occupancy, *i.e.* that all the ribosomes contain a NC. The latter is especially important as the maximum working concentration of the ribosome is limited to 10 µM, which is a concentration that is at least an order of magnitude lower than that is typically used for NMR. A low occupancy therefore, will compromise the signals arising from the NC.

Illustrated in Figure 3.5 is the purification strategy used for *in vivo* RNCs, which was developed during the course of this study (in collaboration with Maria Karyadi, UCL). The strategy includes several key changes relative to the existing purification protocol described in [124], which were found to improve both the yield and purity of the RNCs and these are outlined below.

Following lysis by French press (Figure 3.5 A), the soluble fraction of the lysate is

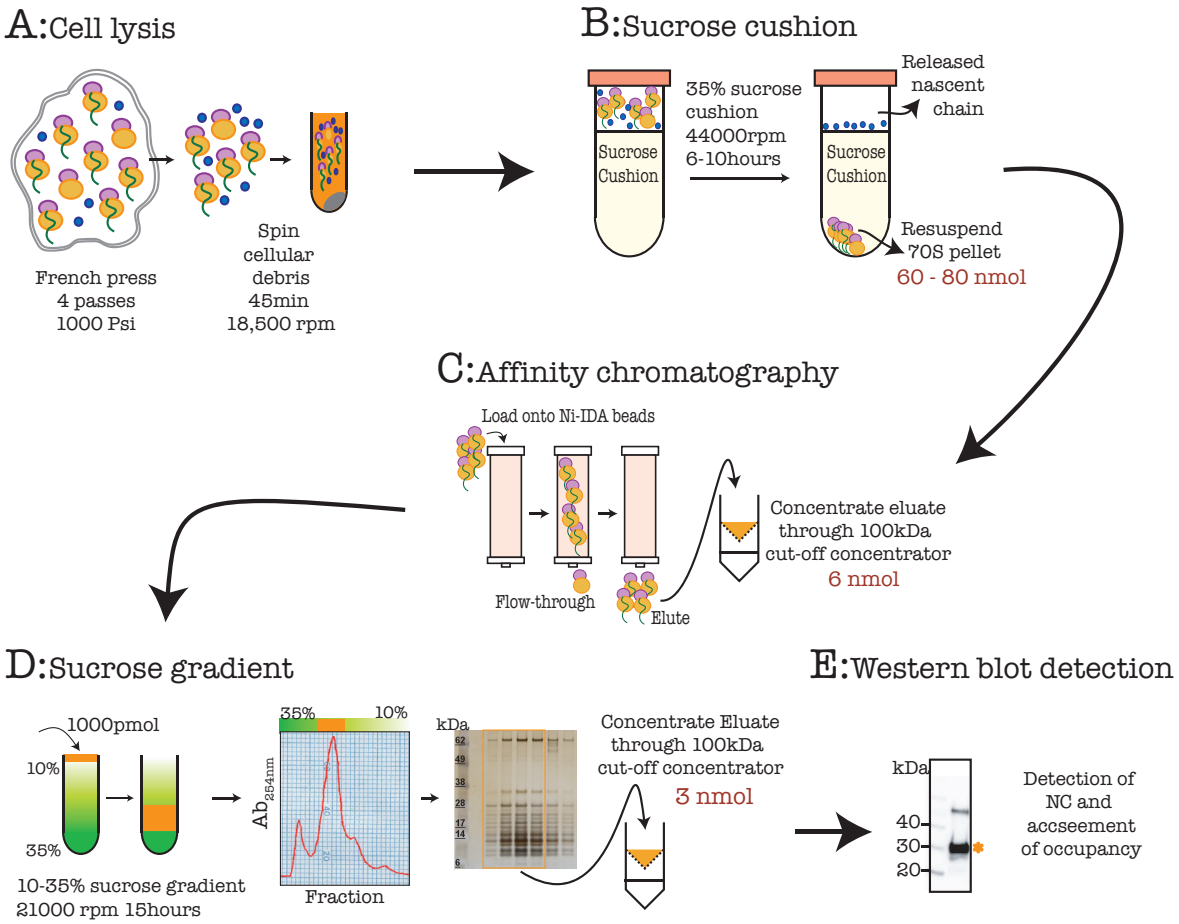

Figure 3.5: Schematic representation of a typical RNC purification following optimisation. **A:** Cell lysis with a French press preserves the ribosomal complexes. **B:** A sucrose cushion separates ribosomal material and large complexes from small soluble proteins. **C:** Metal-affinity chromatography with Ni-IDA resin separates nascent chain-occupied ribosomes from empty ribosomes. **D:** Sucrose gradient followed by fractionation separates intact occupied 70S ribosomes from 50S and 30S ribosomes and factors associated with the ribosome. The purified ribosome-containing fractions identified with silver-stained SDS-PAGE. **E:** A western blot against hexa-his tag assesses the occupancy and overall yield of the RNCs.

subjected to a sucrose cushion (Figure 3.5 B) (prior to metal affinity chromatography), which collects the ribosomal material within the pellet. The introduction of this step (optimised further as described in section 3.2.1.3) was found to improve the recovery of RNCs in the subsequent purification steps by enhancing the purity of the ribosomal material: specifically, the removal of released NCs and other endogenous *E.coli* proteins improved the binding of his-tagged RNCs to the metal affinity column (Figure 3.5 C). Using this approach, the overall yield of Dom5+110-RNCs improved two-fold to 100-200 pmol per litre.

It was found that the type of metal affinity resin was important for improving the

separation of RNCs from empty ribosomes. The TALON resin (cobalt-affinity resin) was replaced with a Ni-IDA resin, as this was reported to have a higher binding capacity than TALON resin (200 mg/ml vs 5-18 mg/ml). The Ni-IDA resin was found to be highly selective for RNCs, and improved the recovery of Dom5+110-RNCs by a factor of 10. The use of the sucrose gradient as described in the original protocol was maintained as this was found to be essential in achieving highly homogenous samples as assessed by SDS-PAGE and western blot analysis (Figure 3.5 D&E). The buffer composition of the sucrose gradient was optimised to improve the overall yield of the RNCs; this is discussed in detail in section 3.2.1.3.

3.2.1.3 Evaluation of buffer conditions for optimising RNC purification

At the beginning of these studies, the average lifetime of RNCs samples was on the order of hours [206]. Sources of RNC sample instability include the degradation of ribosomes due to the release of ribosomal proteins and the release or proteolysis of NCs from the ribosome. Building upon the developments in the expression and purification of the RNCs (described in section 3.2.1.2 above), a set of experiments were performed to evaluate two steps in the purification, the sucrose cushion and the sucrose gradient. This was undertaken in order to improve the purity of the RNCs and therefore extend the lifetime of the samples.

Four buffer conditions were used in this test, the standard Tico buffer (10 mM HEPES, 12 mM $MgCl_2$, 30 mM NH_4Cl, 1 mM BME, pH 7.6), which was used for previous ribosome RNC purifications; 'Buffer B' (50 mM HEPES, 12 mM $Mg(OAc)_2$, 500 mM KOAc, 2 mM BME, pH 7.6), which is a high-salt buffer that has been used to purify RNCs in some studies [126]; and each of these buffers with a supplement of 5 mM ATP, which has been shown to dissociate ATP-dependent co-factors from the ribosome [126]. The four different buffers were tested in different combinations with respect to the cushion, resuspension and gradient steps as shown in Table 3.1.

The cell lysate of Dom5+110 RNC cultures were loaded onto four sucrose cushions made according to the different buffer compositions shown in Table 3.1 (column 2). The resulting

Sample name	Cushion	Resuspension	Gradient
A	Tico	Tico	—
B1	Tico	Tico + 5 mM ATP	Tico
B2	Tico	Tico + 5 mM ATP	Buffer B + 5 mM ATP
C3	Buffer B	Tico + 5 mM ATP	Tico
C4	Buffer B	Tico + 5 mM ATP	Buffer B + 5 mM ATP
D5	Buffer B + 5 mM ATP	Tico	Tico
D6	Buffer B + 5 mM ATP	Tico	Buffer B + 5 mM ATP

Table 3.1: Table of purification tests showing buffer conditions at each of the three purification stages.

pellet from the cushion was then resuspended in either Tico buffer or Tico buffer with a supplement of 5 mM ATP (Table 3.1, column 3). 1000 pmol from each of the resuspended pellets from cushions B, C and D were loaded onto sucrose gradients made in either Tico buffer or Buffer B with a supplement of 5 mM ATP, as shown in Table 3.1 (column 4). In the process, pellets C and D were found to resuspend faster than pellets A and B, which suggests that the high-salt content of the cushion influences the final composition of the ribosomal pellet.

The gradients were fractionated and samples of the fractions were analysed by SDS-PAGE with silver-staining for detailed and high sensitivity detection. Interpretable UV-traces were only obtained from gradients B1, C3 and D5 (as shown in Figure 3.6) because the ATP present in gradients B2, C4 and D6 interfered with the UV-detection. As shown in Figure 3.6, all three traces are similar, but the gel traces show the ribosomal-containing peak shifted towards lower sucrose concentration in B2, C4 and D6, an effect of the high salt concentration. In general, gradients run in Buffer B + ATP show improved purity over those run in Tico, as shown in Figure 3.6, where fewer bands from impurities were observed (ribosomal protein and impurity bands are highlighted with green and red stars, respectively). Therefore, the use of high-salt buffer appears to improve the purity of the RNC sample compared to standard Tico buffer under otherwise identical condition. For a more direct comparison of purity and occupancy, the 70S fraction from the sucrose gradients derived using the novel purification conditions B1 to D6 (Table 3.1) were assessed relative to the original ribosome pellets (A-D) using SDS-PAGE and western blot (Figure 3.7).

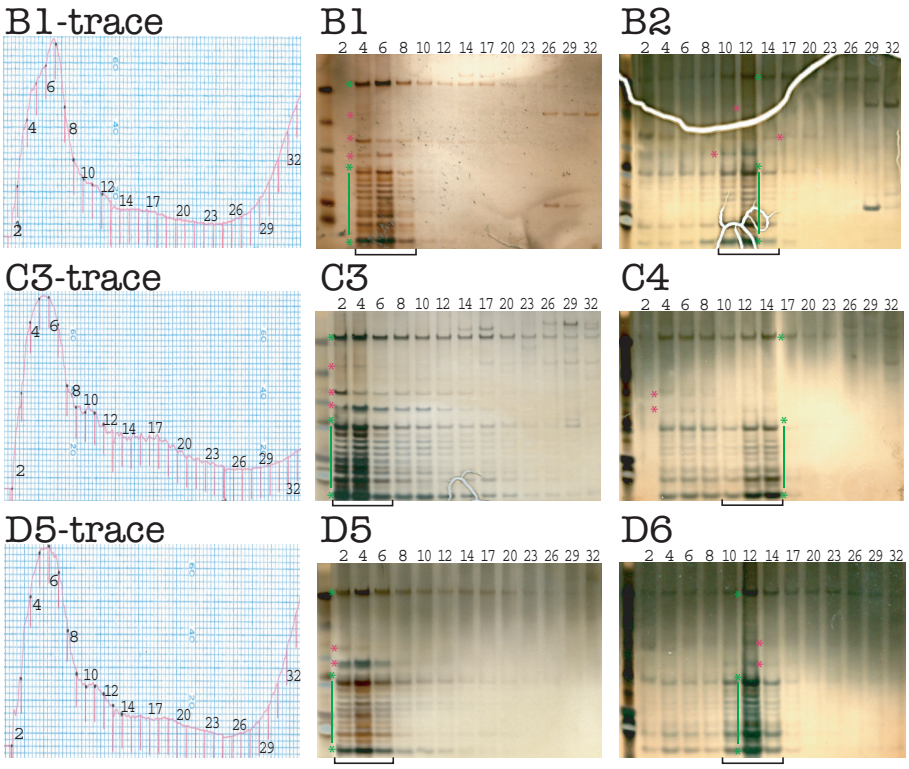

Figure 3.6: Sucrose gradient traces (left column) and silver-stained SDS-PAGE analyses (middle and right column) from the fractionation of sucrose gradients during the test RNC purifications. Numbering of the traces and gels corresponds to the conditions shown in Table 3.1. All gels are loaded from left to right as follows: 5 µl SeeBlue marker, and fractions 2, 4, 6, 8, 10, 12, 14, 17, 20, 23, 26, 29 and 32 collected from the sucrose gradient. The bands highlighted with green stars are ribosomal proteins, and the bands highlighted with red stars are impurities. The intensities of the bands and the colour of the gels are slightly different due to the nature of the silver stain.

The samples of the ribosome pellet that were purified using different buffer compositions (Table 3.1) show the beneficial effect of a high-salt composition buffer, with samples C and D resulting in purer ribosomal material after the cushion compared to samples A and B. The effect of the sucrose gradient was to further improve the level of purity observed in all the samples by removing non-ribosomal contaminants, which typically correspond to bands under SDS-PAGE that fall between *ca.* 40 and 60 kDa. Of the conditions that were evaluated for the sucrose gradient, the most pure samples were obtained in samples D5 and D6 (which were both of comparable purity) (Figure 3.7). Overall, the combined use of a high-salt buffer composition and addition of ATP significantly improved the purity of RNCs compared to the original purification condition represented by sample A (Figure 3.7).

The occupancy of each of the samples was also considered, to ensure that the novel

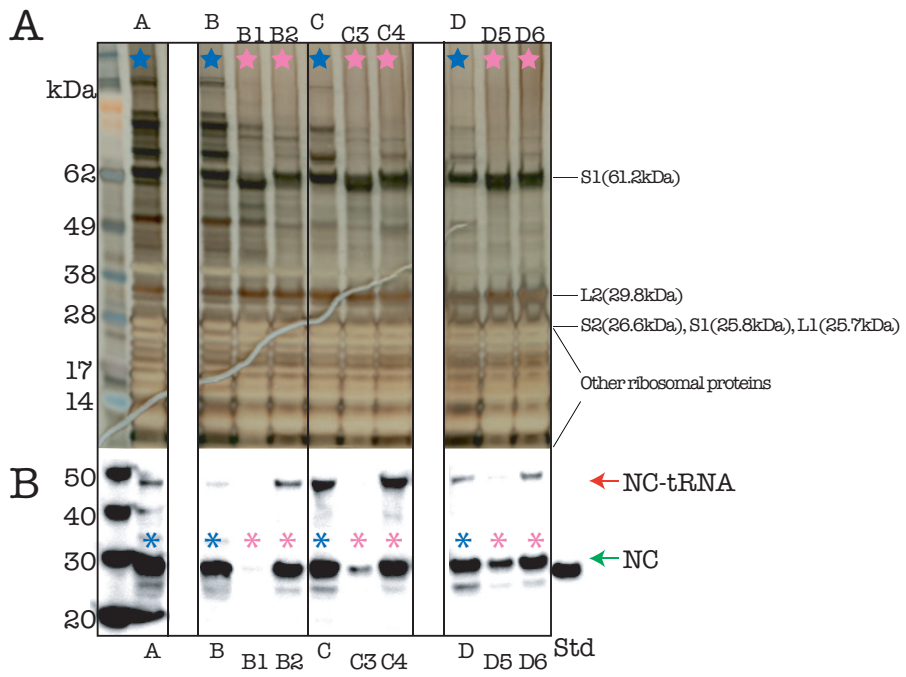

Figure 3.7: Silver-stained SDS-PAGE (A) and anti-his western blot (B) of samples after cushions or gradients. Lanes highlighted with blue stars are samples taken directly after the sucrose cushion, using buffer conditions A, B, C and D from Table 3.1. Lanes highlighted with pink stars are samples after the sucrose gradients, with conditions B1, B2, C3, C4, D5 and D6 from Table 3.1. All lanes were loaded with an equal amount of 10 pmol (ribosomal concentration). Std: protein standard. The NC band (green arrow) is higher than the isolated protein band because of the extra SecM sequence. The higher molecular-weight band (red arrow) is the tRNA-bound form of the NC.

purification conditions did not destabilise the RNC complex. This was evaluated using a western blot (as described in section 5.1.16), which typically visualises RNCs (in this case, Dom5+110) as two dominant bands (Figure 3.7 B, lane A): a 30 kDa band which represents the NC, and a 50 kDa band which is the P-site tRNA-bound form of the NC. It is known that the P-site tRNA can dissociate from the NC at high pH (>8.0) and the relatively alkaline conditions under which the electrophoresis was performed results in partial dissociation of the tRNA from the NC. Lower pH conditions have since been shown to maintain the tRNA-bound form of the NC in Dom5+110-RNCs [242].

In each of the RNC samples, the NC was clearly present within the ribosomal pellet (Figure 3.7, lanes A, B, C and D, green arrow). However there was a variation in the intensity of the NC band following the subsequent purifications using the sucrose gradient. This suggests that during the sucrose gradient, the RNCs are subject to some form of dissociation

Sample name	Yield (pmol)	Estimated occupancy %
A	9658	39
B	15321	31
C	5140	36
D	11541	31
B1	366	—
B2	437	30
C3	534	8
C4	—	29
D5	399	13
D6	422	26

Table 3.2: Table of total yield and occupancy obtained from purification under different conditions. The yield is the total ribosome concentration and the occupancy is determined with the western blot as described in section 5.1.16 on page 196.

or degradation, and the rate at which this occurs is dependent on the buffer conditions.

In Figure 3.7 B, there is much less NC present in lanes B1, C3 and D5, which correspond to Tico-based buffer during the sucrose gradient, than in lanes B2, C4 and D6, where the high-salt + ATP buffer was used for the sucrose gradient stage. When the occupancy was quantified using densitometric analysis, it was found that the Tico-based buffers B1, C3 and D5 gave rise to NC occupancies of 0, 8 and 13%, respectively, compared to the occupancies of the starting ribosomal pellets which were approximately 30%. In contrast, the high salt/ATP-based buffers used for samples B2, C4 and D6 gave rise to occupancies of 30%, 29% and 26%, respectively, which are comparable to those of the starting ribosomal pellets (*ca.* 30%) (Table 3.2).

Together these data suggest that the high-salt/ATP conditions not only improve the purity of the RNC samples, but also better preserve the integrity of the NC, both of which are crucial improvements for the preparation of RNC NMR samples.

3.2.1.4 Conclusions

The investigations of RNC expression and purification yielded significant contributions to the improvement of the RNC production. The expression level was found to be increased both by using EM9 medium and by supplementation with an amino-acid mix. The use of rifampicin reduces the level of background labelling. The purity and integrity of RNCs of Dom5+110

was greatly improved by using high-salt buffer with added ATP. With these improvements, it is now possible to extract 3000 pmol of RNC from 1 l of cell culture, which corresponds to one NMR sample at 10-μM concentration. Moreover, the stability of the RNCs was also improved, with sample lifetimes extended to >24 h.

3.2.2 NMR investigations of the Dom5+110 RNC

3.2.2.1 Preparation of Dom5+110 RNCs for NMR

Using the strategy described in sections 3.2.1 and 5.1.13, Dom5+110 RNCs were generated for NMR. The particular preparation of Dom5+110 described here in which the NC is uniformly ^{15}N labelled corresponds to the NMR analysis discussed in section 3.2.2.2. A 1 l culture was used to express the ^{15}N-labelled Dom5+110 RNC. The OD_{600} was measured to be 5.8 at harvest following induction, demonstrating the high-cell density obtainable using the combination of MDG and EM9 media. Following purification, the sample was of high purity with only a single contaminant band visible by silver-stained SDS-PAGE (Figure 3.8 B, pink star) within the fractions selected from the sucrose gradient (Figure 3.8 A).

After pooling of the relevant fractions (12-20) and exchange into the NMR buffer (10 mM HEPES, 30 mM NH_4Cl, 12 mM $MgCl_2$, 2 mM EDTA, 1 mM BME and 1 μM pepstatin A, pH 7.6), the total yield of RNC was measured as 2900 pmol, corresponding to a 9.8-μM NMR sample.

The occupancy of the RNC and the quantity of bound trigger factor bound were measured via densitometric analysis of the western blot (Figure 3.8 C) according to the method described in section 5.1.16. In brief, the occupancy was determined by quantifying the density of the relevant band in the western blot by integrating the intensity profile over the length of the lane (Figure 3.9 A&B). A standard curve (Figure 3.9 C) was prepared using a range of standards corresponding to varying quantities of isolated Dom5+110. The NC occupancy was then calculated as the ratio of the concentration of detected NC to the ribosomal concentration (measured using UV-spectrophotometer). The NC occupancy was found to be *ca.* 95 %, with trigger factor present in only trace amounts, and below the

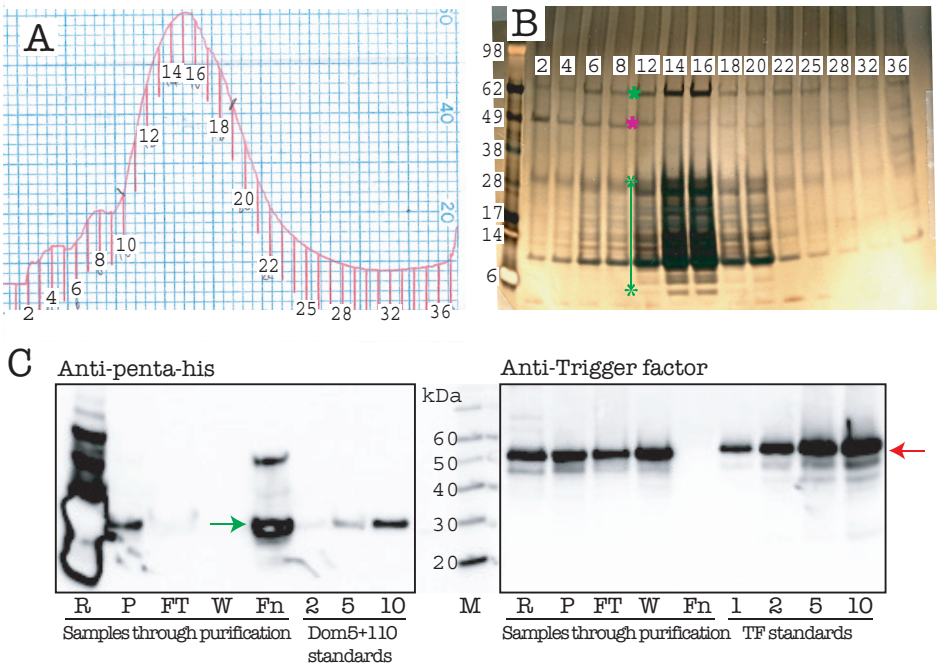

Figure 3.8: **A:** Sucrose gradient fractionation profile and **B:** corresponding silver-stained SDS-PAGE of fractions from (A) of Dom5+110 RNC. Highlighted are the ribosomal proteins (green stars), and the observable contaminant (red star). **C:** Anti-penta-his and anti-trigger factor western blots of the NC during purification. R=released NC, P=pellet of sucrose cushion, FT=flow-through from Ni-IDA column, W=wash of Ni=IDA column, Fn=final sample after sucrose gradient and buffer exchange. The western blot shows that within the final sample (Fn) the overall NC occupancy is ~95 % and quantity of trigger factor is less than 1 %.

detectable limit (less than 1 %).

3.2.2.2 NMR spectra and analysis of Dom5+110 RNCs

Analysis of 2D ^{15}N HSQC spectra of Dom5+110 RNCs. ^{15}N-labelled RNCs of Dom5+110 were produced in *E. coli* (section 3.2.2.1) and 2D ^{15}N HSQC spectra were acquired at 25 °C. The RNC spectrum is shown in Figure 3.10 A. As reported previously [124, 225], the spectra reveal a collection of resonances that are well dispersed, as well as showing a number of heavily overlapped resonances (δ_H=8.8-8.5 ppm) (Figure 3.10 A). When the spectra were compared with that of isolated Dom5, many of the well-dispersed cross peaks of the RNC overlaid closely with the isolated protein (Figure 3.10 B).

Of the observable well-dispersed cross-peaks within the RNC spectrum, 46 peaks (Figure 3.10 D, yellow circles) can be identified unambiguously as signals arising from the folded Dom5 as they overlay with isolated Dom5 peaks to a $\Delta\delta$ value of less than 0.08 ppm,

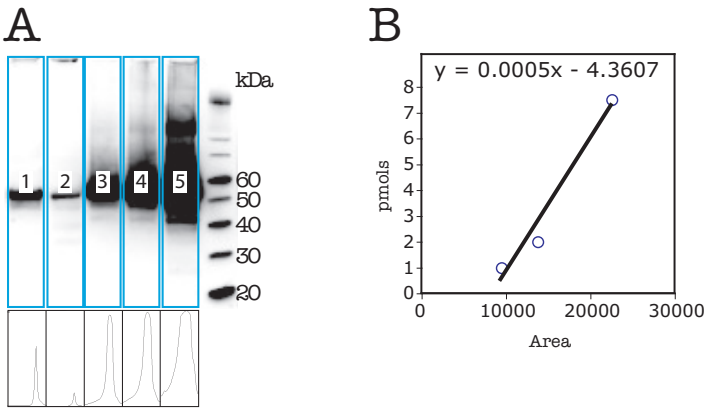

Figure 3.9: Schematic showing densitometric analysis to determine RNC occupancy and ratio of TF to ribosome from western blots (method described in section 5.1.16 on page 196). **A:** Histograms of intensities (bottom) over an area selected for analysis in the western-blot (top, shown here is one RNC sample (lane 1) and standards ranging from 1 to 10 pmol of trigger-factor (lanes 2-5). **B:** A standard curve is obtained by plotting the quantities (in pmol) of the standard samples loaded onto the western blot against the resulting integrated intensities.

where $\Delta\delta = \sqrt{\Delta\delta_{\mathrm{H}}^2 + (\Delta\delta_{\mathrm{N}}/5)^2}$ (Figure 3.10 B).

There are also 31 dispersed peaks in the RNCs spectrum (Figure 3.10 D, green boxes), which do not overlay with the folded Dom5 spectrum but with the L7/L12 resonances in the 70S ribosome spectrum (shown in Figure 3.10 C). This indicates that there is a small degree background labelling of the ribosome during the expression of the RNC. A quantitative labelling experiment is not possible in this case, as both the labelled L7/L12 and the NC would be detected in the 1D ^{15}N-edited spectrum, whereas only the unlabelled L7/L12 would contribute to the ^{15}N-filtered spectrum. The previous investigation discussed in section 3.2.1 suggests a background labelling level of the L7/L12 of ~10 % (Figure 3.4).

Apart from the heavily overlapped region, there is a single cross-peak (black arrow in figures 3.10 B&C) that can not be identified unambiguously because there are cross-peaks at the same position in the spectra of both isolated Dom5 and 70S ribosomes. Hence, this peak was excluded from subsequent analysis of the signal intensities.

The overlapped resonances in the crowded central region (δ_{H}=8.0-8.5 ppm) of the spectrum mostly arise from residues in unstructured or flexible stretches of Dom5+110. This region of the spectrum is difficult to interpret, as many resonances overlay with L7/L12 resonances. However, there are some unfolded resonances that do not overlay with either

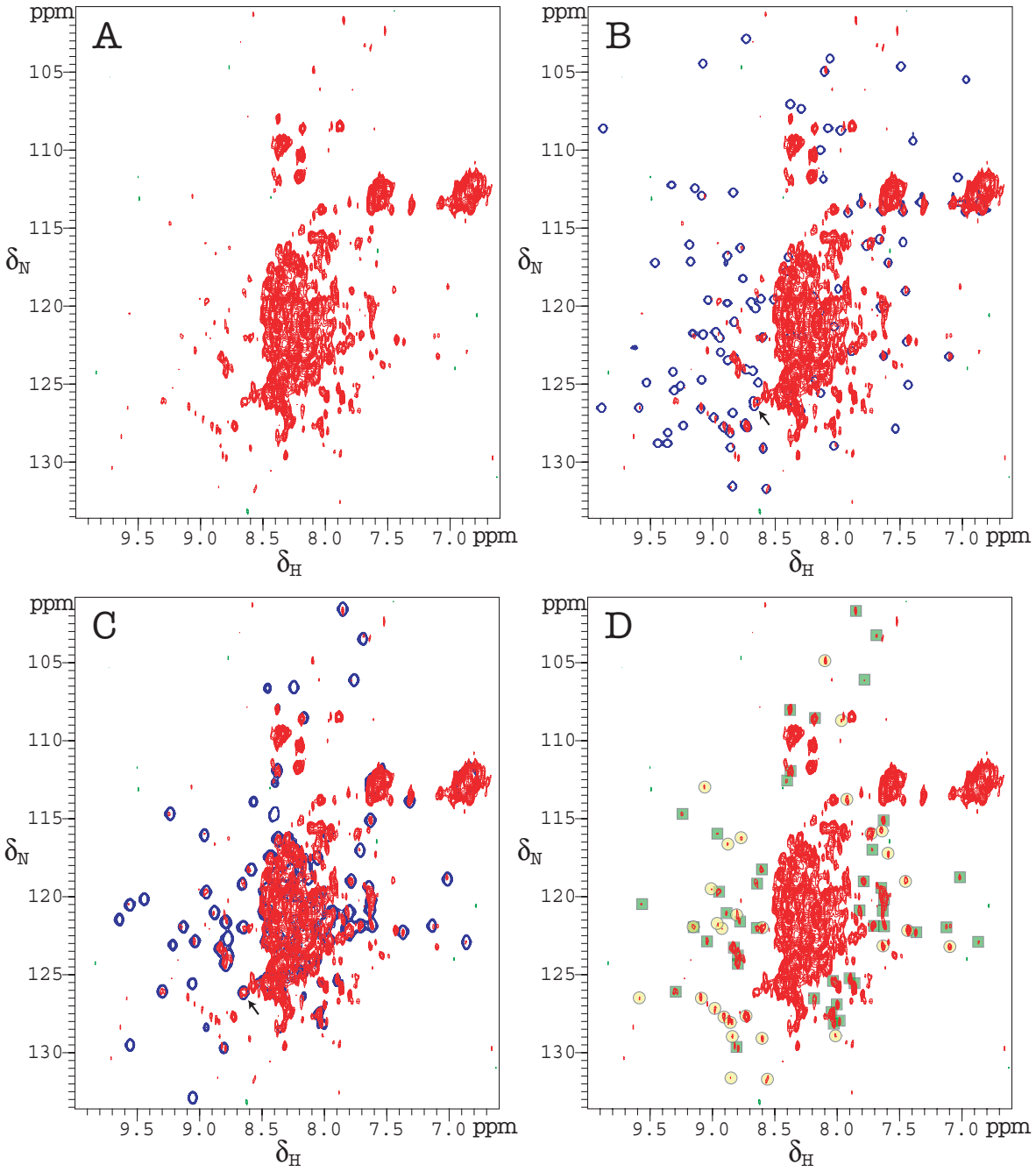

Figure 3.10: ^{15}N-^{1}H HSQC spectrum of Dom5+110 RNC (A) and overlays with the isolated Dom5 protein (B) and 70S ribosome spectrum (C). The isolated protein and ribosome spectra are in blue and the RNC spectrum is in red. The dispersed signals arising from the nascent chain and the ribosome are highlighted in yellow circles and green squares, respectively, in the RNCs spectrum in panel D. The concentration of the RNC is 9.8 μM. All spectra were recorded in Tico buffer (pH 7.6) at 25 °C.

Dom5 or L7/L12 resonances, which most likely arise from the disordered linker region constituted by Dom6 [206]. An assignment of isolated unstructured Dom6 would help to resolve the identity of these resonances, but this was beyond the scope of this study.

Characterisation of the dynamics of ribosome-bound NC via linewidth analysis. The HSQC spectrum of the Dom5+110 RNC was used to measure the linewidth of the observable resonances, providing a preliminary insight into the dynamic properties of the ribosome-bound NC. Overall, the average ^{15}N linewidth of the dispersed resonances from the NC is 22±6.6 Hz, broader than those observed for the isolated Dom5 by a factor of 1.5 (Figure 3.11 A). The broadening observed for the RNCs signals appears to indicate that the relaxation rates of the NC resonances are faster than for those of the isolated protein, which would mean that the motion of the NC is restricted compared to the freely tumbling isolated protein. This is most likely due to the tethering of the NC to the ribosome. The analysis of the linewidths also showed a number of outliers (Figure 3.11 A&B), but these residues have very low signal-to-noise ratios (<4), and hence their estimated linewidths are unreliable and were not considered further.

Although the observed L7/L12 resonances are in principle undesirable, they offer an opportunity to compare the dynamic properties of L7/L12 and the NC. The linewidths of the NC and L7/L12 signals from the same spectrum, within the same sample were similar, with 22±6.6 Hz for the NC and 21±6.5 Hz for L7/L12 (Figure 3.11 B). In addition, the linewidths of the L7/L12 cross-peaks from untranslated ribosomes were found to be very similar to the linewidths of the observable resonances of L7/L12 from the Dom5+110 RNC (within the uncertainty, Figure 3.11 C). As discussed previously in section 2.2.1.7 on page 57, the motional behaviour of L7/L12 CTD on the ribosome is similar to that of an isolated 25-kDa protein. The isolated Dom5 protein is 11 kDa, and therefore the NC appears to tumble at a rate of an isolated protein of twice the molecular weight. This would be consistent with the picture that the linker, while flexible, does not allow the NC to tumble completely independently of the ribosome.

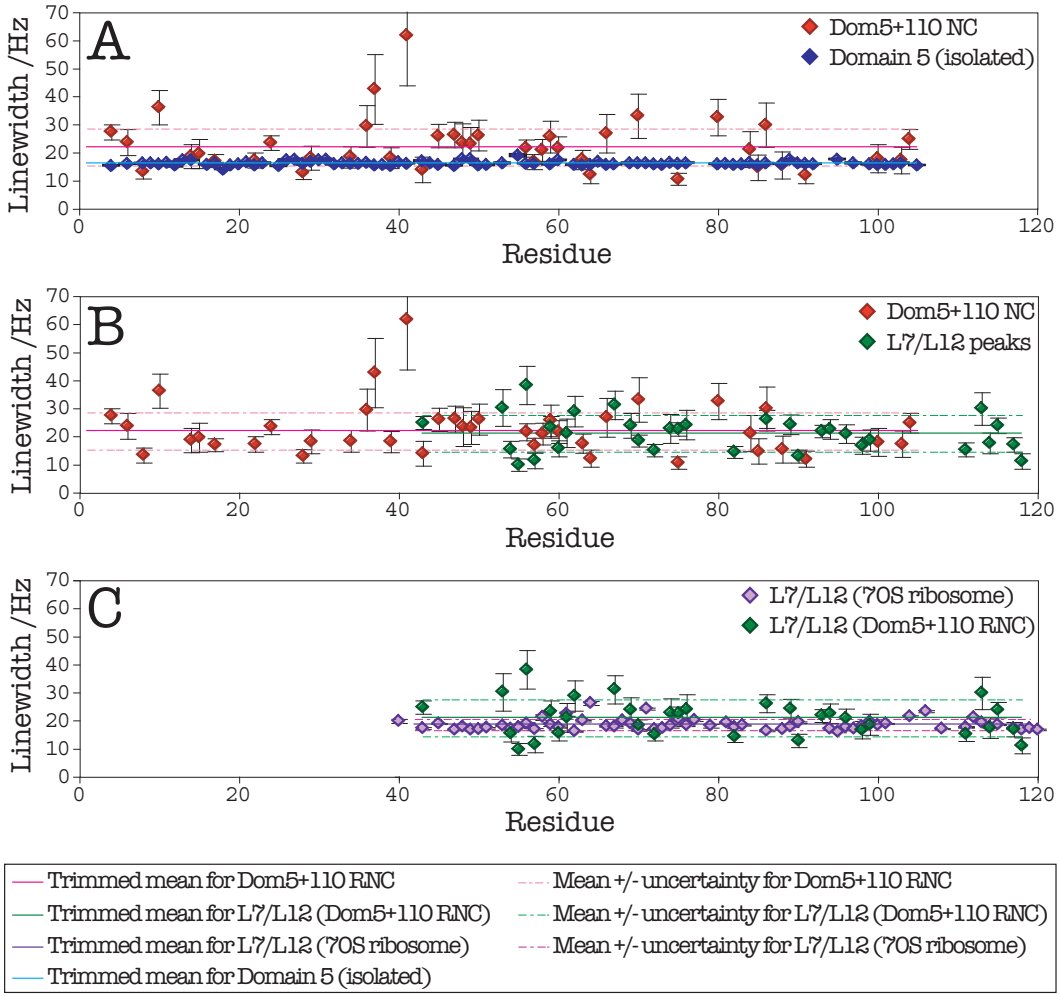

Figure 3.11: A: Plot of linewidths measured for Dom5+110 RNC (red) and isolated Dom5 protein (blue). B: Plot of linewidths measured for Dom5+110 NC (red) and L7/L12 in the Dom5+110 RNC sample (green). C: Plot of linewidths for L7/L12 in the Dom5+110 NC sample (green) and in 70S ribosomes without NC (purple).

Assessing the relative cross-peak intensities of ribosome-bound NC and L7/L12. Given that the relaxation properties of the NC appears to be similar to those of L7/L12, the relative observed intensities should be in proportion to the relative concentrations. The signals arising from L7/L12 on the ribosome are the result of low-level background labelling during expression. The degree of L7/L12 background labelling was measured using an untranslated 70S ribosome sample, which was produced in exactly the same way as the RNCs except that the addition of IPTG was omitted. The efficiency of isotope incorporation into L7/L12 was assessed using the labelling experiments described in section 5.3.1.3 on page 203 (Figure 3.12 A), and it was found that the average background labelling of L7/L12

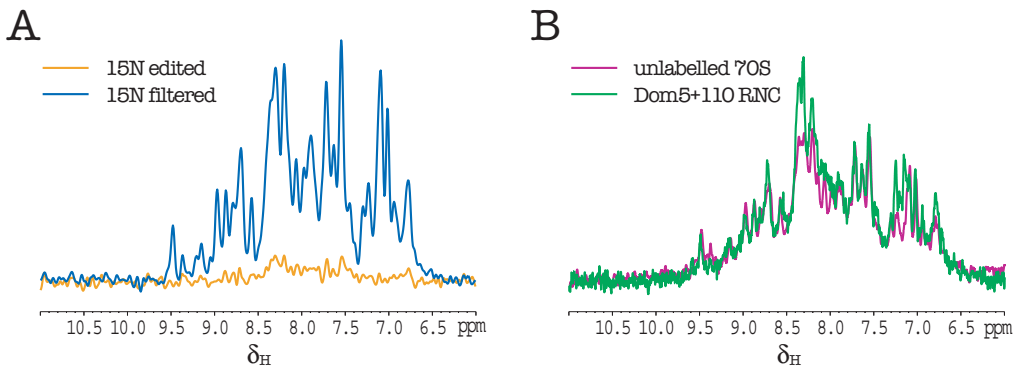

Figure 3.12: **A:** ^{15}N-edited (orange) and ^{15}N-filtered (blue) 1D spectra of 70S ribosomes produced according to the RNC production protocol (section 5.1.13). The ^{15}N-edited spectrum contains signals from protons attached to ^{15}N and the ^{15}N-filtered spectrum contains signals from protons not attached to ^{15}N. The intensity of the ^{15}N-edited spectrum is 10 % of the ^{15}N-filtered spectrum, which means that 11 % of the 70S ribosome N-H groups are labelled with ^{15}N. **B:** Overlays of 1D proton spectra of Dom5+110 RNC (green) and unlabelled, unoccupied 70S ribosomes (magenta).

was 10 %. Since the occupancy of the RNC sample was >98 % (section 3.2.2.1 on page 133), and since two L7/L12 CTDs are visible by NMR, it follows that the concentration ratio of NMR-observable NC and L7/L12 resonances is approximately 5:1 (NC:L7/L12). Therefore, the NC peaks should be five times more intense than the L7/L12 peaks (assuming similar relaxation properties).

Further analysis of the observed NC cross-peak intensities (Figure 3.13 A) showed no sequence-specific differences compared to the intensities of the isolated protein (Figure 3.13 B). However, a comparison of the cross-peak intensities from the NC to those from L7/L12 (Figure 3.13 C) showed that the NC and L7/L12 resonances have very similar intensities (within the error). This finding was supported by the 1D ^{1}H spectra (Figure 3.12 B), in which the envelope of the amide region in the 1D spectrum of the RNC complex overlays closely with that of unlabelled, unoccupied ribosomes, apart from a small region around 8.3 ppm, where the extra signal in the Dom5+110 RNC spectrum can be attributed to the unfolded domain 6 (a more detailed discussion on the spectra of unfolded proteins is given in section 3.2.3). This suggests that the amide proton NMR spectrum is strongly dominated by signals from L7/L12, despite the fact that the NMR-observable proton concentration of L7/L12 is only twice that of the NC.

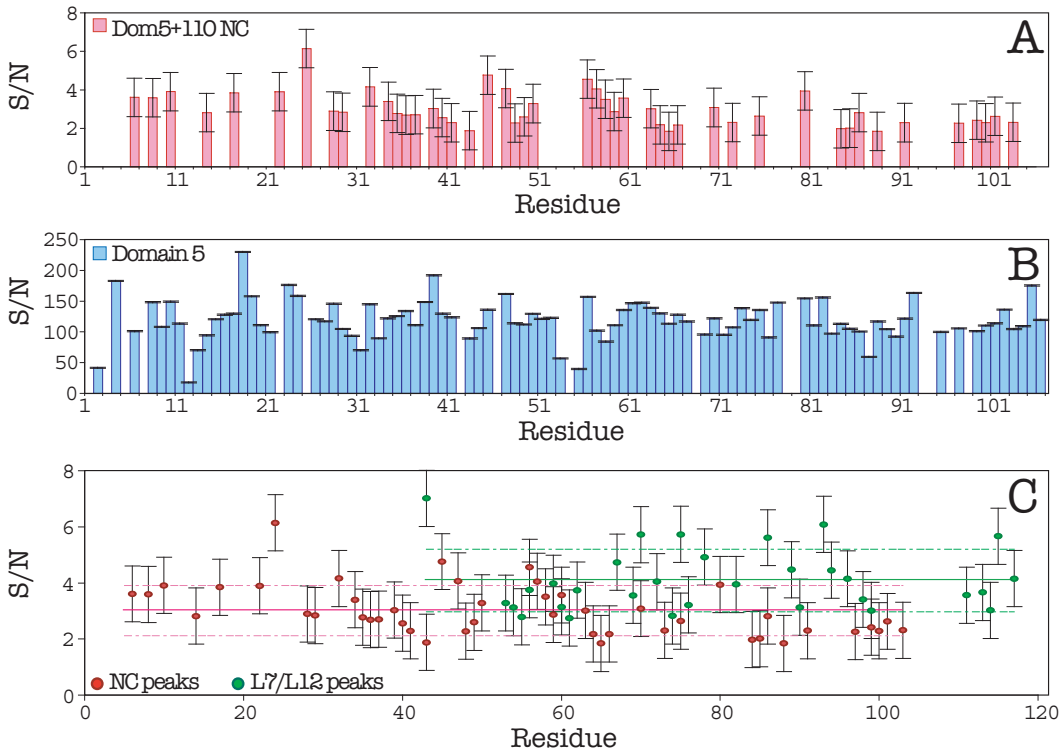

Figure 3.13: A: Plot of signal-to-noise ratio (S/N) of NC peaks in the Dom5+110 RNC spectrum. **B:** Plot of S/N of Dom5 peaks in the isolated protein. **C:** Comparison between the S/N of RNC peaks and L7/L12 peaks in the spectrum of the Dom5+110 RNC. The mean and standard deviation of the NC peaks (3±0.9) are shown with pink solid and dashed lines, respectively. The mean and standard deviation for the L7/L12 peaks (4.1±1.1) are shown with green solid and dashed lines, respectively.

The signal intensities of the NC peaks relative to those of L7/L12 are therefore lower than expected. If this were due to a non-specific interaction between the NC and the ribosome surface, then extensive line-broadening would also be expected, but this is not observed, with the NC peaks having linewidths that are only 1.5-fold broader than those of the isolated protein. In fact, the ^{13}C linewidths measured for the Ileδ1 methyl groups in a U-[^{12}C,^{2}H],Ileδ1-[^{13}C^{1}H] sample of the same NC construct were found to be ~10-fold broader in the NC than in the isolated protein [206]. If the same ratio applied to the ^{15}N linewidths, then the linewidths for the NC would be ~160 Hz, since the average linewidth for the isolated protein is 16 Hz. One explanation for these observations would be that the observed NC signals arise not from the RNC complex, but from a small fraction of released NC. The attachment of the NC was monitored using diffusion experiments, as discussed in the following section.

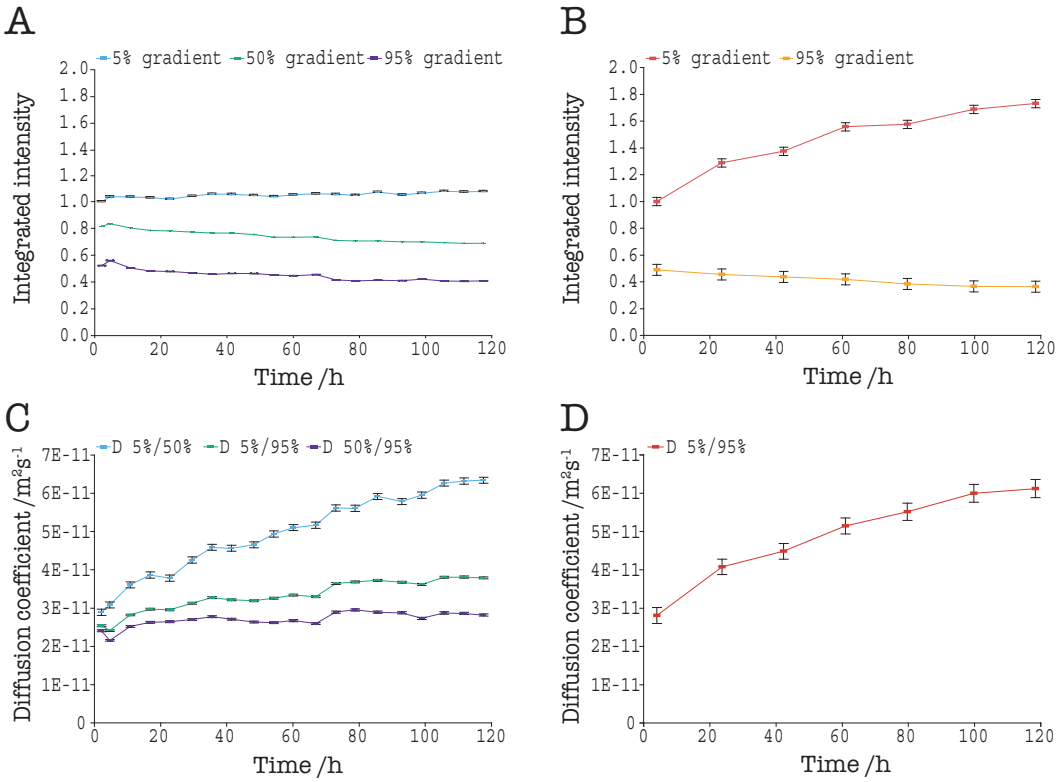

Figure 3.14: Time-course plots of the integrated peak intensities (panels A and B) measured at different gradient strengths (5 %, 50 % and 95 % for ^{1}H (panel A) and 5 % and 95 % for ^{15}N (panel B)), and the extracted diffusion coefficients for ^{1}H (panel C) and ^{15}N (panel D) diffusion experiments (as calculated from the ratio of integrated intensities at different gradient strengths) for the Dom5+110 RNC.

3.2.2.3 Monitoring sample stability with ^{1}H and ^{15}N diffusion experiments

As presented in section 2.2.1 on page 48, diffusion experiments were used previously to monitor the stability of the ribosome and more recently RNCs [207]. A similar approach was applied for RNCs of Dom5+110 to evaluate the integrity of the RNC over the course of the NMR acquisition.

Both ^{1}H and ^{15}N diffusion experiments were recorded for Dom5+110 RNC as described in section 5.3. In this case, the ^{1}H diffusion experiments were used to probe the integrity of the ribosome particle, while the ^{15}N diffusion experiments were used to monitor the nascent chain attachment, which together provide a comprehensive assessment of the RNC.

The integrated intensities at the different gradient strengths and the corresponding diffusion coefficient calculated from the intensity ratios for the ^{1}H and ^{15}N diffusion experiments are shown in Figure 3.14.

From the ^{1}H diffusion experiment (Figure 3.14 A&C), it was seen that the signal intensity for the 5 % gradient (which constitutes contributions from species of all sizes) increased by ~10 % over 120 hours, while the intensity for 95 % gradient (which only has significant contributions from signals due to large particles) decreased by 20 %. This is due to the fact that over time ribosomal proteins dissociate from the ribosome particle, causing an overall increase in intensity in the 5 % gradient strength spectrum and corresponding decrease in intensity in the 95 % gradient strength spectrum. The decrease in intensity for the 95 % gradient (Figure 3.14 A, purple trace) was very small until ~70 hours, at which point a slight step decrease was observed. The intensity then remained almost constant for the remainder of the measurement time. This result is mirrored in the calculated diffusion coefficient derived from the 50 %/95 % ratio (panel C, purple plot), which describes the diffusion of the large particles in the sample, for which the diffusion coefficient remained constant up to 70 hours and then increased, suggesting a slight decrease in the size of the core ribosomal particle at this point.

The ^{15}N diffusion was only measured for two gradient strengths (Figure 3.14 B), 5 % and 95 %, to account for the fact that the ^{15}N diffusion experiment is less sensitive than the simple ^{1}H experiment, and to maximise the measurement time spent recording HSQC and TROSY spectra for RDC analysis. As a result of the low sensitivity for the folded Dom5 resonances discussed in section 3.2.2.2, the observed signal in this instance is mainly from unstructured Dom6. In this case, an increase of the intensity in the 5 % gradient strength spectrum indicates the release of nascent chain, as this species is more dynamic than the ribosome-bound species, and hence gives rise to a more intense signal. The decrease in intensity at 95 % (Figure 3.14 B) was similar to that observed for the ^{1}H diffusion experiment (Figure 3.14 A), indicating that the release of NC occurred very slowly. As only two gradient strengths were used, only a single population and diffusion weighted diffusion coefficient can be calculated. This diffusion coefficient approximately doubles over the course of the measurement, but this is consistent with only a small proportion of released NC. This is because the released NC will contribute much more strongly to the observed signal due to its

slower transverse relaxation as compared to the ribosome-bound NC.

3.2.2.4 Stability of Dom5+110 RNCs monitored by western blotting

Alongside the diffusion experiments and other NMR data that were acquired (section 3.2.2.3), the Dom5+110 RNC sample was subjected to a detailed biochemical characterisation using the same densitometric analysis described in section 3.2.2.1 (Figure 3.15)

In Figure 3.15, the sample before NMR measurement shows two bands: that at ~50 kDa corresponds to the tRNA-bound form of the NC and that at 30 kDa corresponds to the pure polypeptide (lane 1). A sample of the RNC following NMR acquisition for 5 days was loaded in lane 2. As precipitation was occasionally observed for ribosome samples, the post-NMR sample was also subjected to centrifugation to remove the precipitates (lane 3). This did not change the intensity of visible bands (his-tagged molecules), suggesting that the NC was not part of the precipitate. Comparing lanes 2 and 3 to lane 1, it is clear that there is a decrease in intensity for both forms of the original NC, and the occupancy as calculated from the intact NC band is only ~10 % of that in the fresh sample. There is also the appearance of a lower-molecular-weight band (orange arrow), which was suspected to be the N-terminal region of the NC that has been cleaved proteolytically from the ribosome (Dom5, 11 kDa). To identify whether or not the lower band was ribosome bound, the post-NMR sample was filtered through a 100 kDa molecular-weight cut-off concentrator, after which the retentate should contain only the ribosome-bound fraction, while any released/proteolysed species are expected to be in the filtrate. The retentate (green square) and the filtrate (red diamond) were also analysed by SDS-PAGE and western blotting (lanes 4 & 5, respectively), with the resulting bands showing clearly that the lower-molecular-weight species is not ribosome-bound. In addition, no band corresponding to the full-length NC was found in the filtrate, indicating that the NC N-terminal domain was released through proteolysis.

The combination of the diffusion and western blot data indicated that there was some release of the NC over the course of the NMR data acquisition. The western blot analysis suggested that ~90 % of the NC had been released by the end of the NMR acquisition

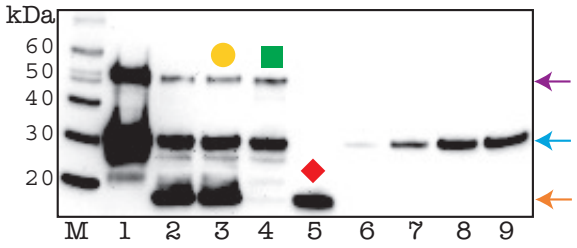

Figure 3.15: Biochemical analysis of Dom5+110 RNC before and after NMR data acquisition: an anti-penta-his western blot of Dom5+110 RNC samples. M=marker, 1=Before NMR, 2=After NMR acquisition, 3=After NMR and spinning down precipitates, 4 & 5=Retentate and filtrate from filtration through a 100 kDa molecular-weight cut-off concentrator, 6-9=Protein standards of 1, 2, 5 &10 pmol, respectively. The purple and blue arrows indicate the position of intact tRNA-bound NC and free NC, respectively. The orange arrow indicates the position of the cleaved product of the NC.

time period (after 120 hours), probably via proteolysis of the unstructured Dom6 linker region. The ^{15}N diffusion data demonstrated that the NC was released gradually, at a relative constant rate, as indicated by the steady increase in the intensity of the 5 % gradient-strength spectrum. Hence, the proportion of attached NC can be estimated at any point during the data acquisition by extrapolation between the fractions bound at the start- and end- points, taken to be 100 % and 10 %, respectively. Based on this approach, together with consideration of the ribosome integrity, data of the first 45 hours of data acquisition were combined for further analysis, at which point ~70 % of the NC remained attached to the ribosome.

3.2.2.5 Approaches to measure N-H RDCs on the Dom5+110-RNC

Following the initial biochemical and NMR characterisation of the Dom5+110-RNC, it was evaluated as a candidate for RDC studies. In the first instance, the isolated protein Dom5 was used to test different methods for partial alignment

Trial with lanthanide-binding tag (LBT) Initially, a lanthanide-binding tag (Figure 3.16 A) was chosen to achieve partial alignment because of the available methods (section 1.4.3 on page 30), and because it is the least perturbing to the protein. The lanthanide-binding site, which is a sequence motif of 17 amino acids, was grafted onto Dom5 (for details see section 5.1.5) in between the his-tag and the N-terminus of the protein. Constructs for both the isolated protein and the RNC complex were prepared. (Figure 3.16 B). The advantage

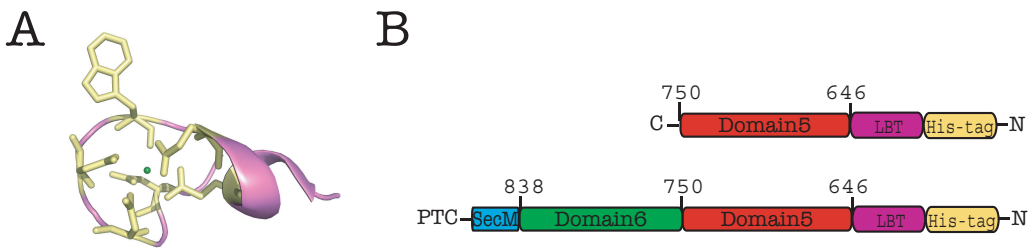

Figure 3.16: Structure of the lanthanide binding tag (A) and isolated-protein and RNC constructs incorporating LBT at the N-terminus of Dom5 (B).

of this method is that it should not interfere with the native structure of the protein [143]. This was confirmed for Dom5, based on analysis of the ^{15}N HSQC spectrum of Dom5-LBT, which overlaid closely with the corresponding spectrum of native Dom5.

Prior to evaluating the LBT for NMR, a fluorescence approach was used to assess the strength of binding between the lanthanide ion and the protein (Dom5). In this experiment, the concentration of the protein was fixed at 1 μM and the ion (in this case Tb^{3+}) concentration was varied between 10 nM and 25 μM. The titration of Tb^{3+} into Dom5-LBT showed a gradual increase in fluorescence but did not reach the plateau expected for saturation of a single binding site. Instead, the data revealed (Figure 3.17 A), a possible plateau at low concentration of Tb^{3+} (red box), which was followed by further increases in fluorescent intensity. This fluorescence profile suggests more than one binding site is available for Tb^{3+}. This hypothesis was confirmed by titrating Tb^{3+} to wild-type Dom5 (Figure 3.17 B), which showed binding of the ion to the protein without the LBT.

Despite the ambiguities observed in fluorescence titrations, RDC experiments were recorded on Dom5-LBT with 0, 0.4, 0.6, 1.0 and 1.5 molar equivalents Tb^{3+}. The observed RDCs for Dom5 were very small (<5 Hz with ~1 Hz uncertainty) as shown in Figure 3.17 C, suggesting that, together with the probable multiple Tb^{3+} binding sites, the lanthanide-binding method for inducing partial alignment was not ideal for RNC studies. Therefore, alternative methods for achieving partial alignment were investigated.

N-H RDCs measured on isolated Dom5. Since the approach of aligning Dom5 with a LBT appeared to be problematic, an alternative alignment medium was investigated, in this

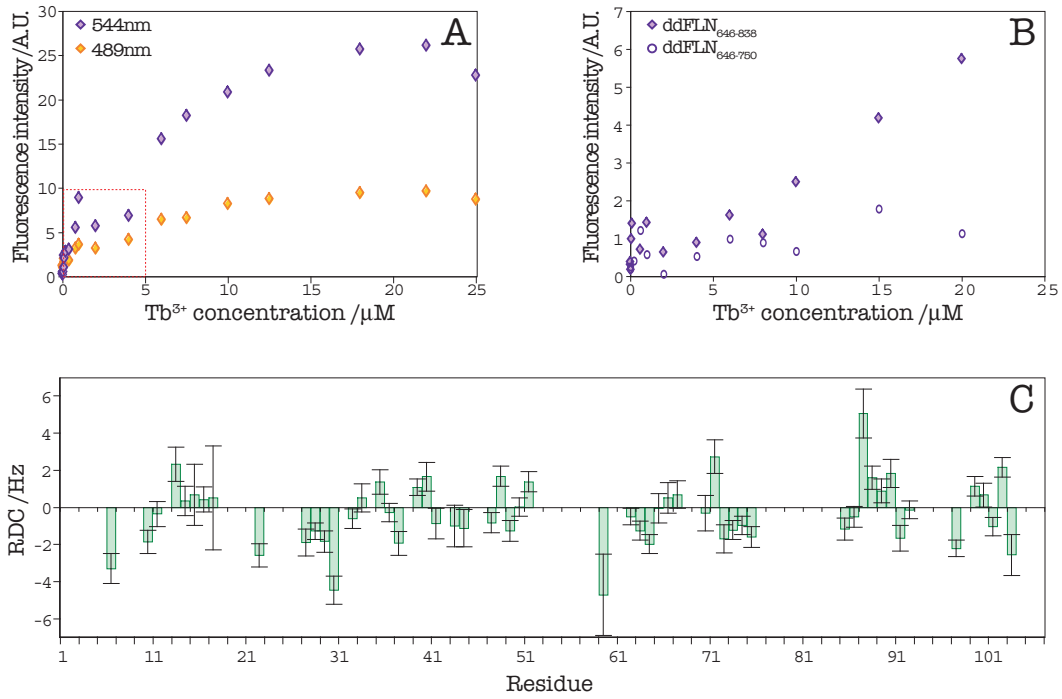

Figure 3.17: A: Terbium titration of 1 µM ddFLN$_{646\text{-}750}$-LBT in 10 mM HEPES, 100 mM NaCl, pH 7.0; λ_{ex}=280 nm, λ_{em}=544 nm/489 nm at 25 °C. **B:** Same as (A) Dom5 and Dom5+6. All points represent independently prepared solutions. **C:** Measured N-H RDCs of Dom5-LBT with 0.6 equivalent Tb^{3+} in 10 mM Hepes, 100 mM NaCl, pH 7.0 at 25 °C.

case Pf1 phage. In contrast to the LBT approach, the phage medium produced a strong alignment of Dom5. The N-H RDCs were recorded on ^{15}N-labelled Dom5 protein using the IPAP-HSQC approach. 81 out of a possible 105 RDCs were extracted, after discarding data from overlapping peaks. The set of 81 RDCs was then trimmed further by selecting only values from residues within rigid secondary structure elements. 43 RDCs from residues in rigid secondary structure were analysed (Figure 3.18 A).

Fitting the RDCs to the crystal structure of Dom5 yielded a Q factor of 0.16, with an RMSD between the measured and back-calculated RDCs of 1.13 Hz (Figure 3.18 B). The RMSD is significantly larger than the RMS uncertainty (0.37 Hz), indicating that the difference between the measured and back-calculated couplings is greater than can be accounted for by the measurement error. This difference is probably due to structural differences between the crystal structure used for the SVD-fitting and the structure in solution, on which the RDCs are measured.

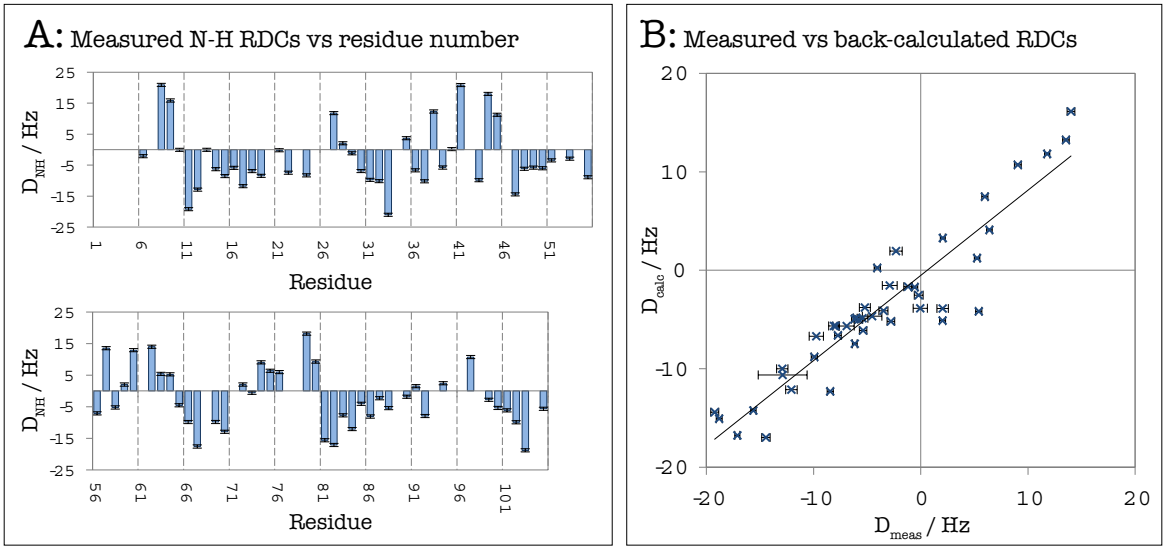

Figure 3.18: **A:** Plot of N-H RDCs measured for Dom5 aligned in phage. **B:** Correlation plot between measured and back-calculated RDCs for Dom5.

Sensitivity of N-H RDC experiments measured on the Dom5+110 RNC For the RNC, the HSQC/TROSY approach for measurement of the N-H RDCs was used instead of the IPAP-HSQC method. The first stage was to measure the isotropic splittings (J couplings) for the Dom5+110 RNC. The quality of its isotropic ^{15}N HSQC and TROSY spectra was similar to the spectrum shown in Figure 3.10. The signal-to-noise ratio for the dispersed peaks was low and only 13 residues of folded Dom5 were unambiguously assignable in these spectra. The J couplings for these 13 residues were calculated and the associated uncertainties estimated according to their signal-to-noise ratios and measured linewidths (Equation 2.4). The average uncertainty for the J-couplings was 20.4 Hz, which is clearly too large to give useful data when considered in relation to the expected range of RDCs (magnitude 0-25 Hz). In addition, the linewidth analysis described in section 3.2.2.2 suggests that the observable peaks in the HSQC spectra that appear to belong to folded Dom5 may arise from small quantities of released protein, rather than from the ribosome-bound NC.

In conclusion, further work is required to improve the quality of the RNC spectra before a detailed RDC study can be undertaken.

3.2.3 NMR investigations of α-synuclein (αSyn) RNC

As Dom5+110 was found to be a challenging candidate for RDCs, an alternative system, α-synuclein, was also examined, as it represents an ideal model system for NMR studies of intrinsically disordered protein, and gives intense and resolvable resonances within its ^{15}N HSQC spectrum.

αSyn, together with β and γ-synuclein, belongs to the synuclein family of proteins, which are expressed throughout the body but express at particularly high levels in the brain, most abundantly in the pre-synaptic nerve terminals [243, 244]. αSyn is an intrinsically disordered protein (IDP) of unknown function strongly implicated in Parkinson's disease (PD), which is a degenerative disorder of the central nervous system. αSyn self-associates into amyloid fibrils, contributing to the formation of inclusion bodies known as Lewy bodies [245].

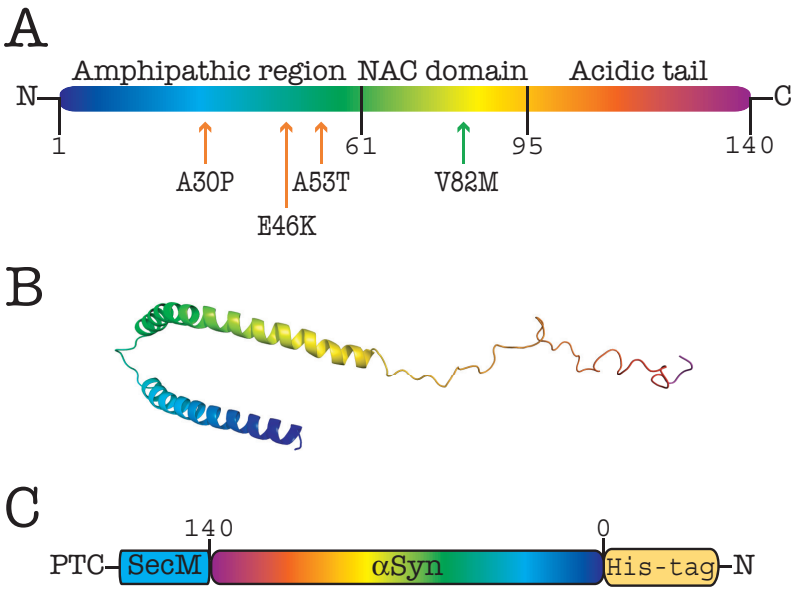

Figure 3.19: A: αSyn can be divided into three regions, the N-terminal amphipathic region, the self-aggregating hydrophobic NAC domain and the negatively charged and proline richC-terminal tail. The disease related mutations are highlighted with orange arrows and the V82M mutation point is labelled with a green arrow. B: The part-helical structure of αSyn upon binding to bicelles (PDB:1XQ8) [246]. C: Schematic for the αSyn-RNC construct.

The αSyn sequence can be divided into an N-terminal amphipathic region (residues 1-61), the non-β-amyloid-component (NAC) domain (residues 61-95) and the acidic C-terminal tail (residues 95-140) (see Figure 3.19 A). The N-terminal amphipathic region

contains the three sites of mis-sense mutations (A30P, E46K and A53T) that are involved in familial forms of early-onset PD by increasing the rate of amyloid fibril formation, and together with the NAC domain, contains seven imperfect repeats (XKTKEGVXXXX) including the conserved KTKEGV motif. The central NAC domain contains the hydrophobic stretch VTGVTAVAQKTV, which can self-associate and has been shown to be essential for aggregation and fibrillation [247]. Although it is an IDP, the overall structures of αSyn is more compact than that of a completely random coil protein [248]. NMR studies show that it has an average of *ca.*10% α-helical content in the N-terminus, which increases upon binding to membranes (see Figure 3.19 B) [246]. The C-terminal region of αSyn is negatively charged and has been shown by NMR to be involved in long-range interactions with the central hydrophobic region [168, 248].

To understand the progressive biosynthesis of an IDP and probe its conformational preferences on the ribosome, RNC complexes of αSyn were prepared and used to evaluate RDCs as a technique for describing its structural properties.

3.2.3.1 Preparation and stability of α-synuclein (αSyn) RNCs for NMR

To explore the possibility of applying the RDC approach to the study of αSyn on the ribosome, SecM-stalled RNCs of ^{15}N-labelled αSyn were generated as described in section 3.2.3.1. Preliminary studies of αSyn RNCs identified an unexpected single point mutation (V82M), which enhanced the stability of the the resulting RNC (in collaboration with Annika Weise, UCL). To examine in more detail the NMR characteristics of αSyn RNCs, both WT and V82M variants were subjected to further analysis.

Preparation of the sample. Two αSyn-RNC constructs were used, the wild-type (WT-RNC) and a V82M variant (V82M-RNC). The procedure for expressing and purifying both RNCs constructs was as described in section 5.1.13. The isolated proteins of both constructs were also prepared as described in section 5.1.11. It was found that there were significant differences in the characteristics of the two RNC constructs. For example, the total ribosome content recovered after the sucrose cushion was found to be higher for WT-RNCs, but most

of these were unoccupied and passed through the metal-affinity column without binding. Similarly, the material in the elute of the column was typically greater for V82M-RNC than for the WT-RNC (Table 3.3). Overall, the final yield of the V82M-RNC was also higher than that of the WT-RNCs by approximately 30 %. It is unclear why the V82M mutation improves the recovery of the αsyn-RNC, but it is likely to be due to an enhancement in the stability of the RNC complex.

Purification stage	WT-RNC /pmol	V82M-RNC /pmol
Pellet after cushion	186960	144460
Flow-through from column	166800	77184
Elution of column	3840	6804
Final sample after gradient	1392	2100

Table 3.3: Summary of αSyn-RNC purification: shown is the total ribosomal material (in pmol), for WT- and V82M-RNCs during purification. Both RNCs were expressed under identical conditions.

During the purification process, the WT-αSyn-RNCs gave rise to a fractionation profile after the sucrose gradient that contained two distinct peaks, an early peak (pk1) and a later peak (pk2) (Figure 3.20 A). The two peaks were indistinguishable by SDS-PAGE, and it is possible that pk1 represents a higher order assembly of ribosomes (*e.g.* disomes or polysomes); both fractions (pk1 and pk2) were evaluated separately (see below).

Using the western blot shown in Figure 3.20, the occupancies of the samples were determined according to the method described in section 5.1.16 on page 196. In panel B, which is the anti-αSyn western blot of the WT-RNC, in which the total amounts of ribosomes loaded in all lanes were the same (apart from the standards), the band corresponding to the NC (green arrow) is most intense in sample 'pk2' (blue box, Figure 3.20 A), which is the sample used in NMR experiments. The estimated occupancy for 'pk2' was calculated to be 98 %. Panel C is an anti-trigger factor (TF) western blot, from which the ratio of TF to ribosome was determined using similar method as for determination of the RNC occupancy. The calculated ratio was 3.8 %, which is sufficiently low to avoid significant effects on the NMR observability of the NC. A similar western blot analysis for the V82M-RNC is shown in Figure 3.20 D&E. The final sample 'Fn' again shows the most intense band (greater

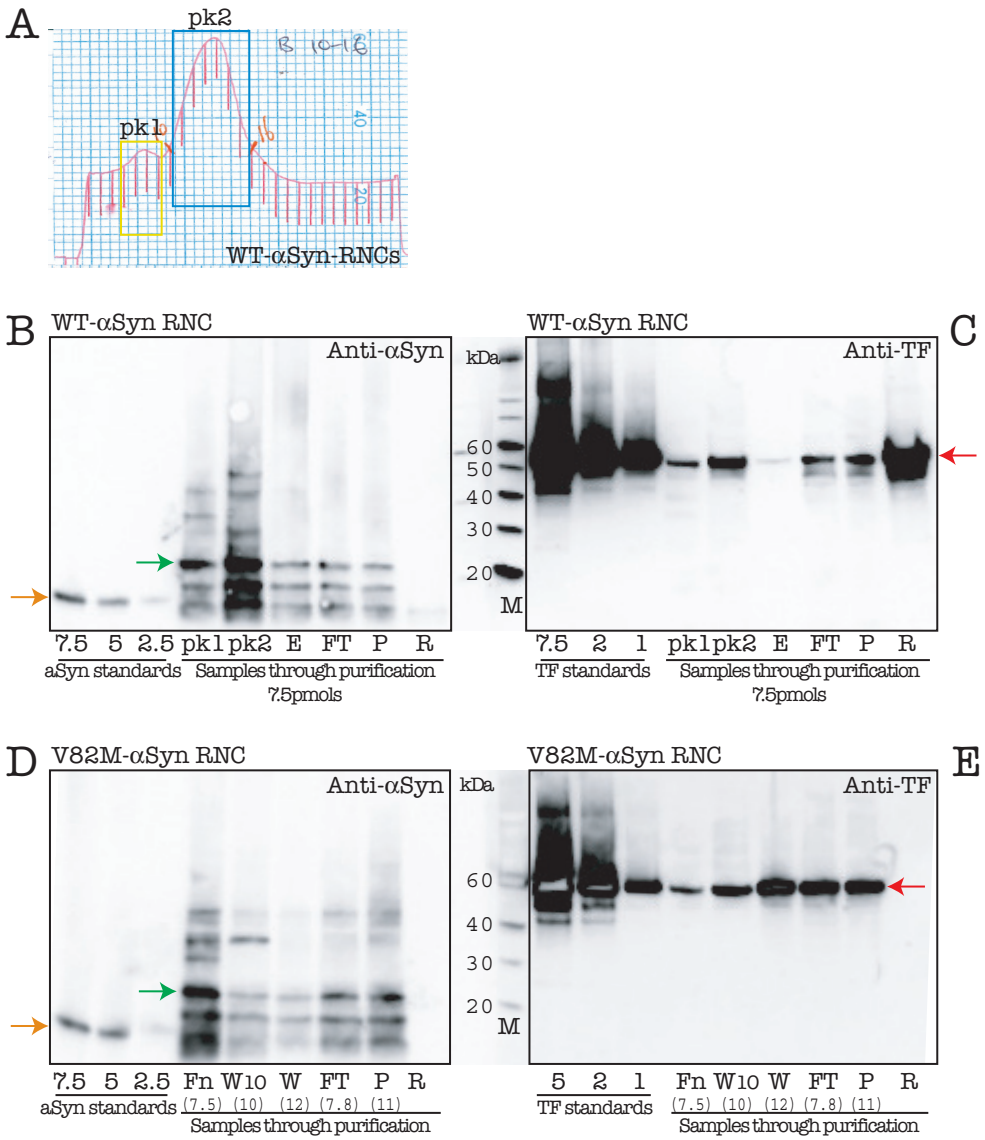

Figure 3.20: A: Sucrose gradient fractionation profile of WT-αSyn RNC, showing two peaks, pk1 and pk2. **B-E:** Anti-trigger factor (C&E) and anti-αSyn (B&D) western blots of WT-RNC (B&C) and the V82M-RNC (D&E) for samples collected at different stages of the purification. The bands corresponding to trigger factor and αSyn from the RNC samples are highlighted with red and green arrows, respectively. The bands corresponding to the αSyn protein standards are indicated by orange arrows; the αSyn standards migrate faster than the NCs because the αSyn protein strandards do not contain the SecM stalling sequence present in the NC construct. M=marker, R=released NC, P=pellet after sucrose cushion, FT=Ni-IDA column flow-through, E=column elute, pk1=first peak in the gradient trace (yellow box, panel A), pk2=second peak in the gradient trace (blue box, panel A), W=column wash, W10=column wash with 10 mM imidazole, Fn=final sample. Numbers in brackets indicate the amount of sample for SDS-PAGE in pmol. The standards are labelled with their amounts in pmol.

than 98 % occupancy) compared to the other bands corresponding to the different stages of purification (panel D). The anti-TF western blot in panel E, in contrast, shows the least intense band in the final sample 'Fn', and the ratio of TF to ribosome is calculated to be 2 %. This estimated ratio was calculated by comparing the intensity of the band from the final RNC sample to the 1 pmol band of the trigger factor standard. It would have been more accurate to plot a standard curve derived from all three standards, but this was not possible here as the more concentrated TF standards were over-exposed.

Monitoring the stability of αSyn-RNCs by western blotting. To evaluate the integrity of the WT- and V82-RNC samples during RDC experiments, a biochemical approach was implemented in addition to the NMR-based monitoring. For both WT- and V82M-RNCs, an aliquot taken from the fresh NMR sample was incubated separately at 4 °C and samples were extracted at different time points over a period of ~170 h. These were centrifuged through a sucrose cushion to separate the ribosome-bound and released NCs, which are then localised to the pellet and the top-of-cushion fractions, respectively. All the samples were examined by anti-α-synuclein western blotting in which the intensities of the top-of-cushion (T) and pellet fractions (P) were compared to uncentrifuged samples (A) to quantify the population of ribosome-bound NC.

For the WT-RNC after 24 hours of data acquisition, the αSyn band appears exclusively in the pellet fraction, indicating that the NC was fully ribosome-bound. To test this conclusion further, a control was also performed in which released NC was added to 70S ribosomes at an equivalent concentration and stoichiometry (1:1) to the RNC samples taken during NMR. The control revealed the presence of the NC in the top-of-cushion fraction alone (Figure 3.21 A, lane '-ve'), confirming that the sucrose cushion assay was suitable for discriminating between released and ribosome-bound species.

The quantities of the pellet fraction could not be accurately determined as a proportion of the starting concentration of material (7.5 pmol). Nevertheless, the consistent observation of the αSyn species solely in the pellet fraction demonstrates that the nascent chain remains ribosome-bound for up to 166 hours at 4 °C.

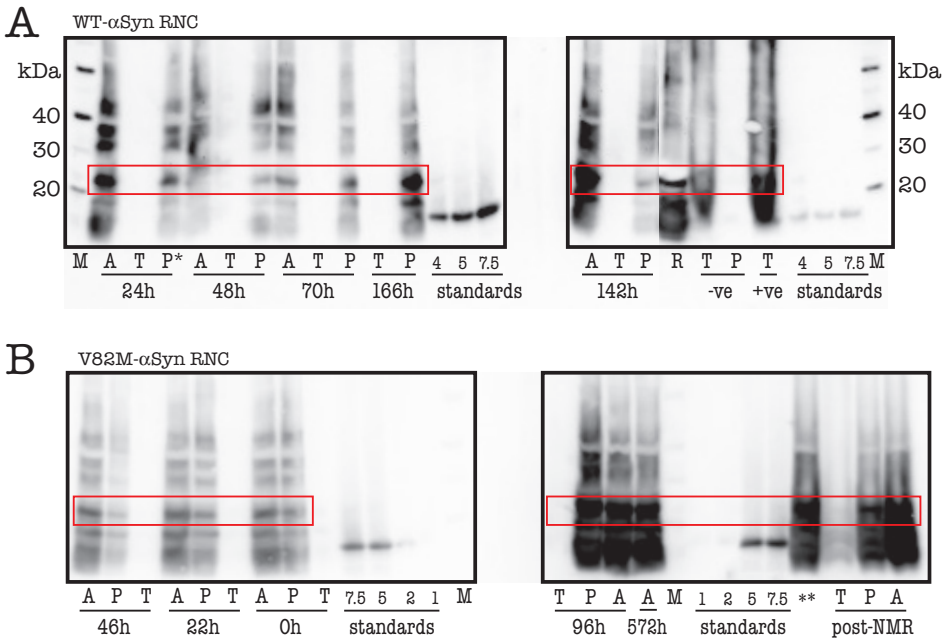

Figure 3.21: Anti-αSyn western blot of samples taken during NMR data acquisition of **A:** WT-RNCs and **B:** V82M-RNCs. The position of bands corresponding to the NC is highlighted with the red box. **A:** M=marker, A=whole sample (7.5 pmol), T=top-of-cushion (released/cleaved NC), P=pellet of cushion (attached NC, 7 pmol), R=purified V82M released NC, -ve=negative control (mixture of 120 pmol of 70S ribosomes and purified V82M released protein loaded onto the cushion, +ve=positive control (purified V82M protein loaded onto the cushion), *=3 pmol loaded onto gel. **B:** M=marker, A=whole sample (7.5 pmol), T=top-of-cushion (released/cleaved NC), P=pellet of cushion (attached NC), **=sample before addition of phage (after buffer-exchange to remove released/cleaved NC, 7.5 pmol). The RNC lanes are labelled with the time in hours at which the sample was taken. The protein standards lanes are labelled according to the quantity loaded (in pmol).

A similar result was observed with V82M-RNCs, which were used in subsequent RDC experiments (section 3.2.3.4). Over the first 96 hours, during which isotropic HSQC and TROSY spectra were recorded (*i.e.* in the absence of Pf1 phage), the RNC sample was observed to be stable with the NC observed only as a ribosome-bound species. At 96 hours, Pf1 phage was added to the sample and aligned spectra were recorded (section 3.2.3.4). A sample taken after these sets of experiments (Figure 3.21 B) shows that the nascent chain remains ribosome-bound, despite the extended incubation period (250 hours) and the presence of phage. The extended lifetime of the αSyn-RNC compared to Dom5+110 RNC is at least partly due to the low temperature (4 °C) at which the αSyn-RNC NMR experiments were recorded.

3.2.3.2 NMR analysis of αSyn RNCs

NMR spectra for both αSyn isolated protein and RNC samples were recorded at 4 °C. Recording at a low temperature slows the rate of solvent exchange for amides in disordered proteins, yielding sharper peaks in the spectra [249]. It also advantageous for RNCs because the reduced temperature extends the life-time of the samples.

The 2D ^{15}N HSQC spectrum of the isolated WT αSyn protein is shown in Figure 3.22 A. Its characteristics are typical of an unfolded protein and agree well with previous spectra of αSyn [65]. It has very narrow ^{1}H chemical shift dispersion (7.6 - 8.7 ppm) compared to a typical folded protein (6.5 - 10 ppm). While the $^{1}H^{N}$ chemical shift range typically becomes more dispersed in the presence of secondary or tertiary structure, the ^{15}N chemical shift is mostly dependent on the observed residue-types and the residue-types of the flanking residues, so that the ^{15}N chemical shift range is highly dispersed even for disordered proteins.

The ^{15}N HSQC spectrum of the isolated V82M protein (Figure 3.22 D), was found to be similar to the isolated WT protein spectrum (Figure 3.22 A), with a small number of chemical shift changes, which arise from residues close to the mutation point (V82M). The backbone chemical shift assignment of V82M was performed by Dr Chris Waudby (UCL), and these assignments have been used in this study for the analysis of the RNC spectra.

The ^{15}N HSQC spectra of WT- and V82M-RNCs are shown in Figure 3.22 B&E. The ^{1}H chemical shift dispersion for both NCs remains narrow and the WT-RNC spectra overlay well with the corresponding isolated protein spectra (panel C&F), indicating that the NCs are unfolded on the ribosome. Obtaining accurate peak counts for the RNC spectra is difficult due to the intrinsically overlapping nature of the αSyn spectrum, where all the peaks are crowded into the unfolded region. However, it is clear that the V82M-RNC spectrum contains fewer peaks than the WT-RNC spectrum. 83 and 34 peaks were assigned unambiguously in the WT- and V82M-RNC spectra, respectively. The overall decrease in peak-count might be a result of overall line broadening in the RNC spectra. The lower number of assigned peaks in the V82M-RNC spectrum is partly due to the lower overall peak counts, but is also due to increased overlap that prevents transfer of assignments from

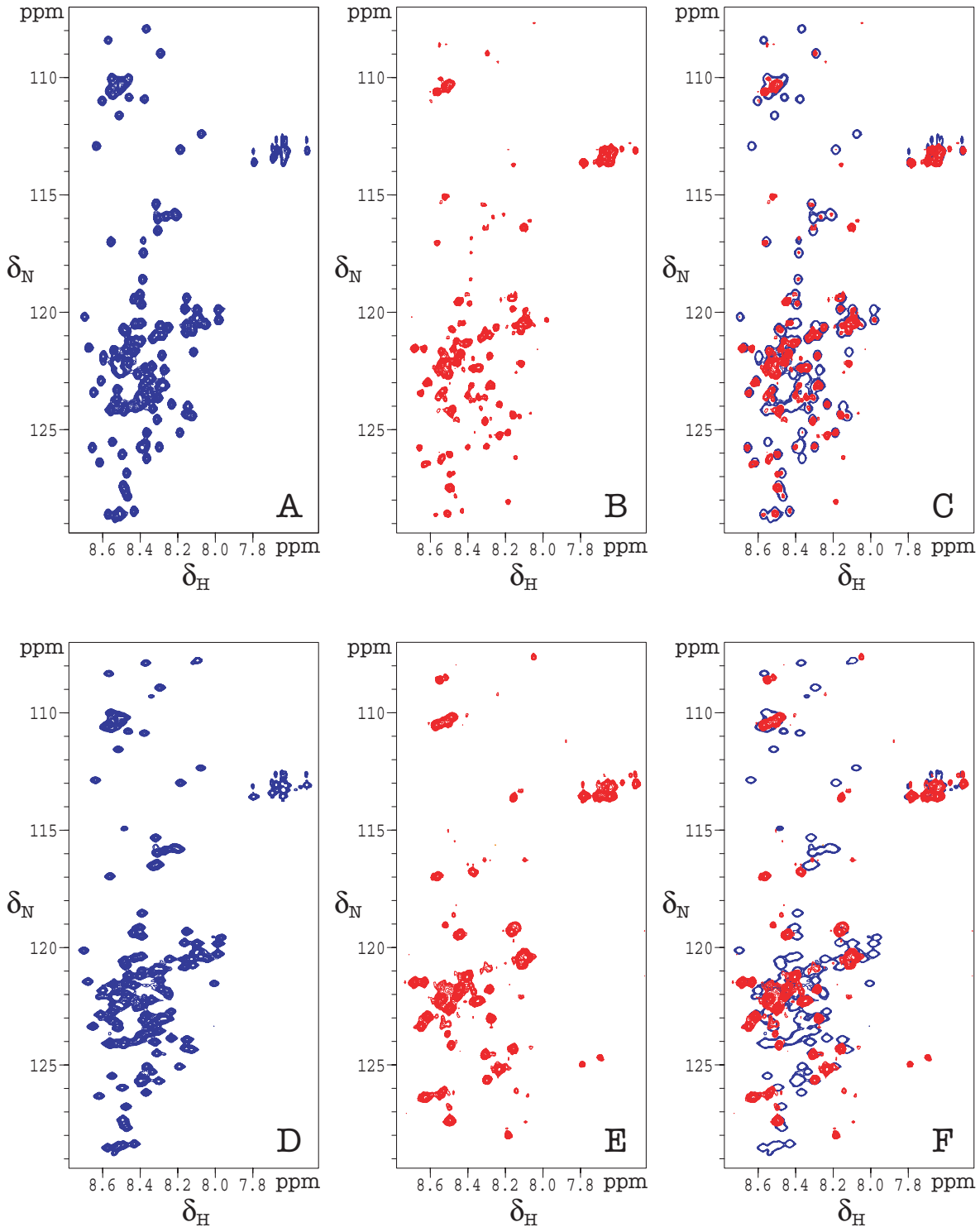

Figure 3.22: ^{15}N HSQC spectra of WT αSyn (A-C) and V82M αSyn (D-F). The isolated proteins are in blue and the RNCs are in red. Panels C and F are overlays of the RNC spectra with the isolated protein spectra (the isolated protein spectra are shown with fewer contour levels for clarity). The concentrations of the WT-RNC and V82M-RNC are 4.3 μM and 5.25 μM, respectively. All spectra were recorded in Tico buffer at pH 7.6 at 4 °C.

the isolated protein spectrum. The latter difficulty is illustrated in Figure 3.23, which shows a peak cluster in the isolated and V82M-RNC spectra. While 8 peaks are assigned for this cluster in the spectrum of the isolated protein (panel A), none of these assignments can be transferred to the spectrum of the RNC, despite five local maxima being detectable (panel B, black crosses).

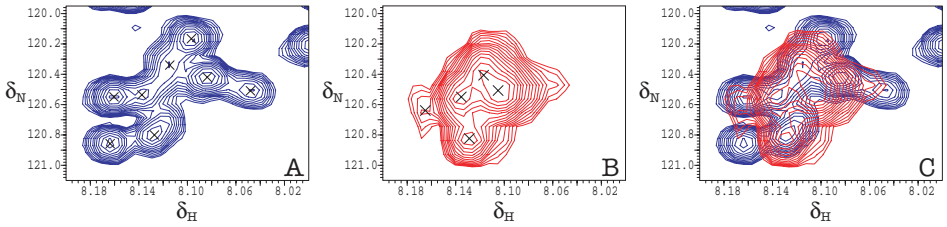

Figure 3.23: Selected region of spectra from isolated V82M protein (A) and V82M-RNC (B), illustrating the increased peak overlap as a result of broadening. C: Overlay of spectra from panels A and B. The black crosses indicate local maxima that are detected by the auto peak-picking routine.

To investigate if the increased overlap and loss of peaks in the RNC spectra were due to line-broadening, the ^{15}N linewidths of the assignable peaks in the RNC spectra were measured and compared with those of the corresponding peaks in the isolated proteins. For WT αSyn, the average linewidths (calculated over a set of ~30 peaks) were 11.8 Hz and 6.7 Hz for the RNC and isolated protein, respectively. For V82M, the average linewidths (calculated over a set of ~20 peaks) were 8.6 Hz and 7.4 Hz for the RNC and isolated protein, respectively. Therefore, the RNC signals are broader than those from the isolated proteins, and the broadening appears more severe for the WT.

This comparative linewidth analysis is complicated by the fact that the measured linewidths for the RNCs are only from peaks that appear with sufficient intensity in the spectra. This means that the average linewidths calculated over the sub-set of observed peaks for the RNC spectra do not represent the true average linewidths over all residues in the NCs, but are skewed averages, which are biased towards the narrower, and hence more observable, peaks. An additional problem is that many of the observable peaks in the RNC spectra cannot be assigned due to overlap and small chemical shift changes, again reducing the number of peaks for which the linewidths can be reliably compared between the isolated

protein and RNC spectra.

The first problem is demonstrated by the average linewidths for the WT and V82M-RNC spectra. The average linewidth is lower for the V82M-RNC, but this is calculated over fewer peaks than for the WT-RNC. Many of the peaks that are very broad in the WT-RNC spectrum are completely absent in the V82M-RNC spectrum, suggesting the true average linewidth over all residues is larger for the V82M-RNC. However, it is clear that both RNCs show broader lines than their corresponding isolated proteins, indicating that the relaxation properties of the ribosome-anchored αSyn protein are different from those of free αSyn. One possible explanation for this difference is that it arises as a result of tethering the protein to the ribosome, which therefore restricts its conformational flexibility. However, this effect is expected to be small because the dynamics of intrinsically disordered proteins are dominated by local motions, which should not be affected significantly by anchoring one end of the chain.

To further explore the differential dynamics between the isolated proteins and their RNCs, the ^{15}N HSQC intensity profiles (along the sequence) were plotted for each of the four αSyn samples (Figure 3.24). Firstly, these profiles highlight that there are missing peaks in the RNC spectra compared to those from the isolated proteins, particularly for the V82M construct. Secondly, it is apparent that the majority of missing peaks for the RNCs are clustered in the N-terminal half of the sequence. For the WT-RNC, where more peaks are observed, those peaks that are still detectable in the N-terminal half of the sequence are in general weaker than those towards the C-terminus. This picture does not agree with the hypothesis that the attachment of the NC to the ribosome is restricting its flexibility. If this were the case, then the C-terminus of the NC would be less flexible than the N-terminus. The observed intensities suggest instead that it is the N-terminus whose motion is restricted.

For the V82M-RNC, one explanation consistent with the observations is that the N-terminal portion of the protein has been degraded by proteolysis. However, this explanation cannot be applied to the WT-RNC, where peaks are observed as close to the N-terminus as residue 10. Another explanation is that there is non-specific interaction between the N-

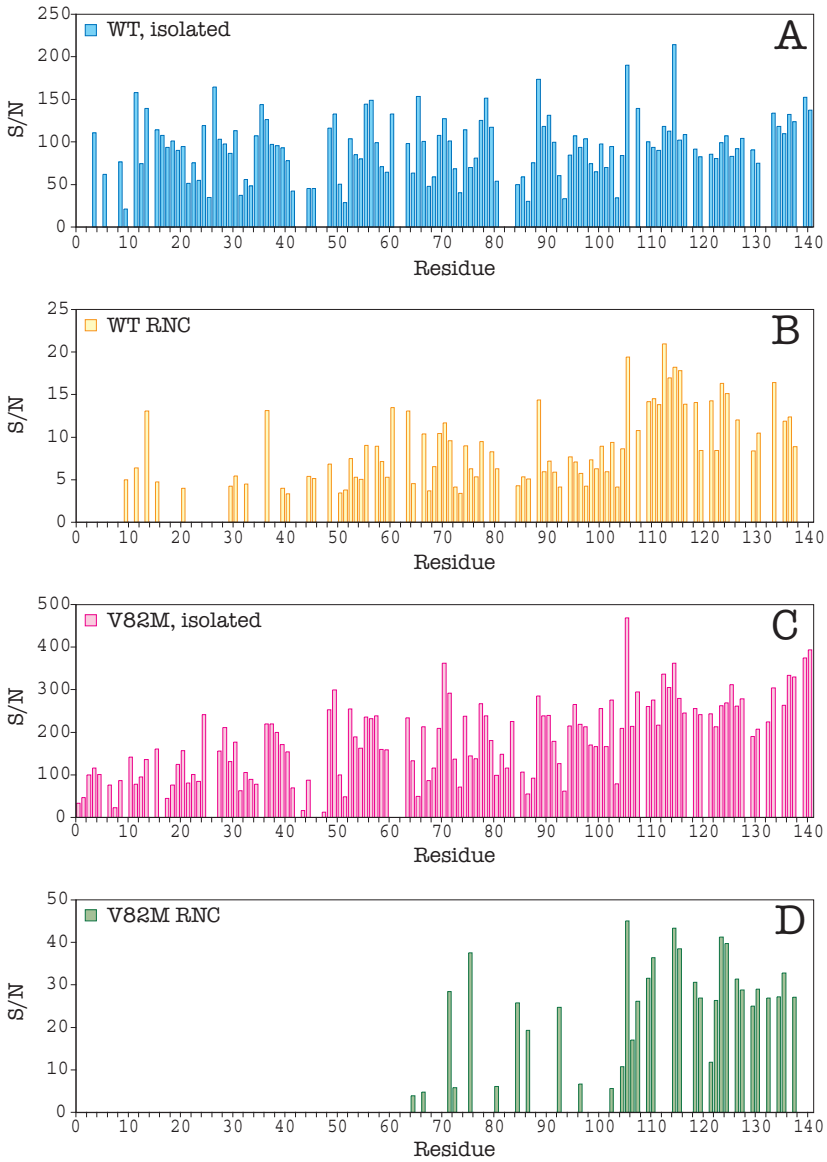

Figure 3.24: Plots of signal-to-noise ratios (S/N) against residue number for isolated WT αSyn (A), the WT-RNC (B), isolated V82M (C) and the V82M-RNC (D).

terminal part of the αSyn NC and the ribosome. These hypotheses were investigated further and are discussed later in section 3.2.3.5.

3.2.3.3 Monitoring RNC sample stability using ^{1}H and ^{15}N diffusion experiments

Before considering the RDC data for analysis, diffusion experiments and western blots (section 3.2.3.1) were used to ensure that the recorded NMR data reflected attached NCs. ^{15}N diffusion experiments were recorded throughout the course of NMR data acquisition for both the WT- and V82M-RNCs. In the case of the V82M-RNC, the isotropic splittings

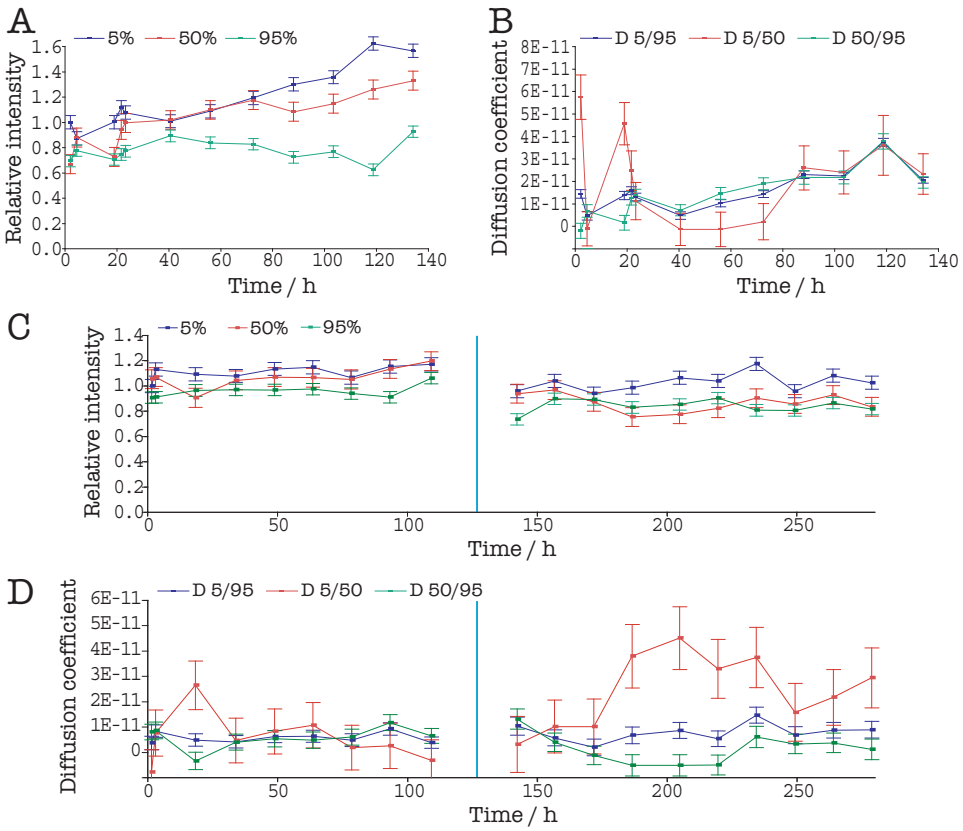

Figure 3.25: Plot of the integrated peak intensities (panels A and C) and the calculated diffusion coefficients (panels B and D) from the ^{15}N diffusion experiments recorded on the WT-RNC (panel A and B) and the V82M-RNC (panel C and D). The blue bar in panels C and D indicates the point at which phage were added to the V82M-RNC sample.

were recorded first, after which phage were added and the aligned splittings measured. For the WT-RNC, only the isotropic splittings were recorded. Plots of the integrated signal intensities in each diffusion row and the diffusion coefficients calculated from the pair-wise intensity ratios are presented in Figure 3.25.

Overall, the signal intensities and calculated diffusion coefficients are relatively stable for both RNCs. Comparing the WT-RNC with the isotropic stage for the V82M-RNC, the diffusion coefficients for the V82M-RNC are stable and homogeneous up to 110 hours, while those for the WT-RNC begin to gradually increase after 40 hours. This suggests that the V82M-RNC appears to be slightly more stable than the WT-RNC, as also observed during their purifications (section 3.2.3.1).

After addition of phage to the V82M-RNC, there was a slight decrease in the intensity of all rows of the diffusion experiment (after re-scaling to account for the decreased

concentration). This small decrease is probably due to magnetisation losses during the pulse sequence caused by evolution of proton-proton dipolar couplings, and mismatches of the aligned N-H splittings to the INEPT delays. While the intensities remained relatively stable for the remainder of the measurement, there was an apparent increase in the diffusion coefficient calculated from the 5 % and 50 % gradient strengths, possibly indicating some release of NC. However, further inspection indicates that the diffusion coefficient calculated from the 50 and 95 % gradient strengths decreases over the same time period, which is implausible. The explanation is that the intensity for the 50 % gradient strength was systematically under-estimated for these time-points: the relative intensity plot shows that 50 % gradient strength intensity is lower than that for the 95 % gradient strength for this period. This demonstrates the challenges involved in acquiring high-quality diffusion data on RNC samples. Due to the low concentration of the RNCs, the diffusion spectra are noisy, leading to relatively large uncertainties in the calculated diffusion coefficients. The random uncertainty can be estimated (section 5.4.1), but it is more difficult to account for systematic errors. In particular, small baseline offsets that vary between the rows can influence the calculated diffusion coefficients. Careful baseline correction was undertaken for the data presented here, but poor water suppression caused by a damaged Shigemi tube means that small offsets may still be present. A small negative baseline offset is the most likely explanation for the underestimation of the intensity in the 50 % row.

Overall, the diffusion data confirm the biochemical analysis, demonstrating the NCs remain predominantly ribosome-bound over the extended period of NMR data acquisition (up to 270 hours on the V82M-RNC).

3.2.3.4 Approach to measure N-H RDCs on αSyn-RNC

Prior to measuring RDCs on the αSyn-RNCs, the RDCs of the isolated V82M protein was investigated, following the procedure outlined in section 2.2.3.1, by which the isolated V82M protein was aligned in Pf1 phage. In this case, the N-H RDCs were measured using the HSQC/TROSY method, to maintain a consistent approach with the RNCs. The measured

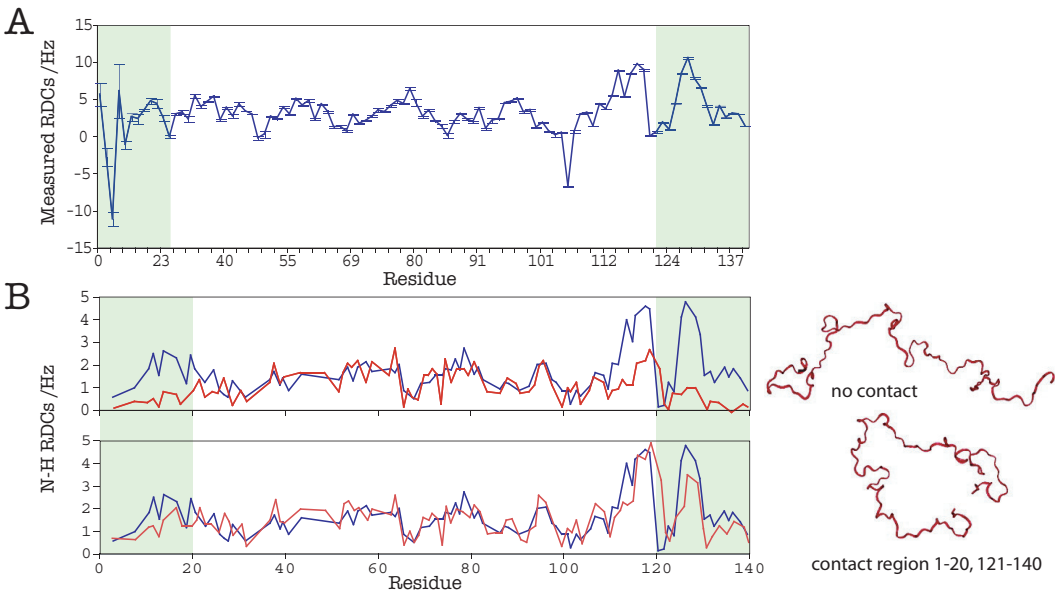

Figure 3.26: A: Plot of measured N-H RDCs for isolated V82M protein. **B:** Plot of measured (blue) and simulated data (red). The green shading indicates regions where simulated data do not match the measured data if there are assumed to be no long-range contacts. Panel B generated using data taken from [168].

RDCs are plotted along the sequence in Figure 3.26. The RDCs measured on the isolated protein (Figure 3.26 A) show a similar trend to the published data [168]. The regions highlighted with green shading were shown by simulation to make long-range contacts (Figure 3.26 B). The region between residues 120 and 140 also shows higher RDCs than the rest of the sequence, which could be indicative of residual local structure.

A similar approach was then adopted for the corresponding RNCs. Before proceeding with extended data acquisition, a set of relatively short (2.5 hours) experiments were undertaken with the aim of assessing which experiment type would be optimal for measuring N-H splittings of intrinsically disordered NCs. NCs consisting of folded domains are similar to the case of ribosome-bound L7/L12, for which the HSQC/TROSY approach is clearly preferred over the IPAP-HSQC method. However, for disordered NCs, the preferred method is not necessarily the same as that for folded NCs. This is because disordered NCs are expected to have significant local conformational flexibility, and hence slower relaxation rates in general. Therefore, the differential relaxation between the upfield and downfield components of the ^{15}N doublets is expected to be less severe than for folded NCs or ribosome-bound L7/L12, so that the IPAP technique may be preferred over the

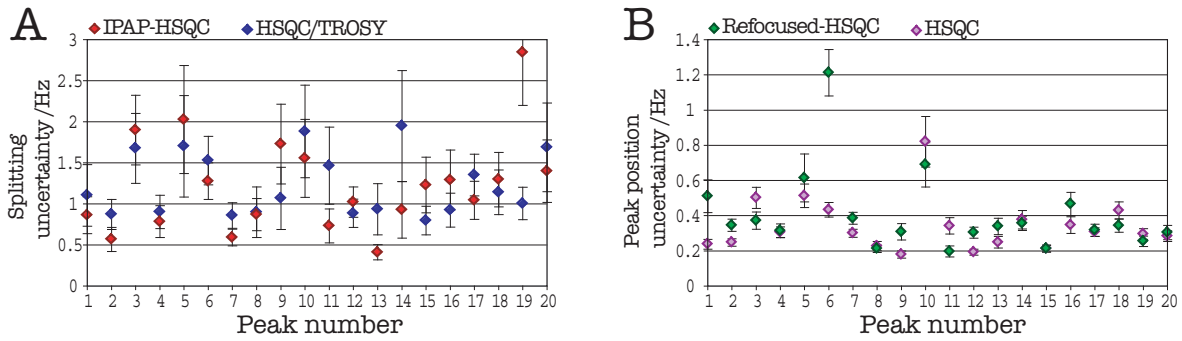

Figure 3.27: **A:** Plot of splitting uncertainties calculated for HSQC/TROSY and IPAP-HSQC methods. **B:** Plot of peak position uncertainties calculated for standard HSQC and refocused HSQC sequences.

HSQC/TROSY approach.

Therefore, two comparisons were made. First, the HSQC/TROSY spectra were compared to the IPAP-HSQC sub-spectra. Second, a refocused ^{15}N HSQC sequence with semi-constant time ^{15}N evolution and ^{1}H composite-pulse decoupling [250] was compared to the standard HSQC sequence. The refocused ^{15}N HSQC sequence should yield narrower ^{15}N linewidths, thereby improving the precision at which the peak positions can be defined. For the first comparison, the estimated uncertainties in the extracted splittings were calculated for a set of 20 well-resolved peaks based on their signal-to-noise ratios and linewidths. The uncertainties are calculated for the full N-H splitting in both approaches, *i.e.* the uncertainties in the peak position differences from the HSQC/TROSY spectra are doubled to obtain the uncertainties in the N-H splitting. For the second comparison, the peak position uncertainties from the standard and refocused-HSQC spectra were calculated for the same set of 20 peaks.

The calculated uncertainties are presented in Figure 3.27. It was found that there was little difference between the IPAP-HSQC (average uncertainty 1.22±0.58 Hz) and HSQC/TROSY (average uncertainty 1.23±0.38 Hz) approaches. In addition, the refocused HSQC sequence was not found to offer significant benefits over the standard sequence. While the ^{15}N linewidths were found to be slightly narrower (the average linewidth decreased by 0.24 Hz) using the refocused sequence, there was a concomitant drop in signal intensity that outweighed the decrease in linewidth, resulting in a higher peak position uncertainty than for the standard sequence (0.4±0.2 vs 0.34±0.15 Hz). The remaining data were therefore

recorded using the HSQC/TROSY approach with the standard HSQC sequence. This approach was deemed slightly advantageous compared to the IPAP method, despite the similarity in the projected uncertainties, because the combined resolution in the HSQC and TROSY spectra is still expected to be better than in the two sub-spectra obtained from the IPAP/HSQC. This consideration could be important in the context of the overlapped spectrum of the unfolded NC.

Isotropic splittings were measured initially on the V82M-αSyn RNC, yielding the spectra shown in Figure 3.28 A. Following this, the sample was buffer exchanged and concentrated to a smaller volume for addition of phage. The deuterium splitting after addition of phage was 13.8 Hz.

The HSQC spectra before and after addition of phage were similar (Figure 3.28 A&B) with only small chemical shift changes of less than 0.05 ppm (Figure 3.28 C). No noticeable line-broadening was observed, with the changing by less than 2 Hz. This suggests that there was no significant interaction between the phage and the NC and that the relaxation properties of the NC were not affected by the addition of phage.

24 RDCs were measured for V82M-αSyn RNC, predominantly from residues at the C-terminal end of the sequence. The measured RDCs vary between -5 and +10 Hz, with an uncertainty of 1 to 2 Hz, *i.e.* 10-20 % of the largest RDC measured. This could potentially be improved by increasing the sample concentration to 10 μM, which would double the signal-to-noise ratio and therefore halve the uncertainty.

The current set of RDCs are too incomplete to be used for further structural analysis. An incomplete set of RDCs can still provide useful information for a folded protein because they can be used to derive the alignment tensor. However, the data analysis techniques for unfolded proteins are more demanding and require at least one relative complete set of RDCs to provide meaningful results. Further work is therefore required to establish the underlying cause for the loss of N-terminal signals in the αSyn RNCs.

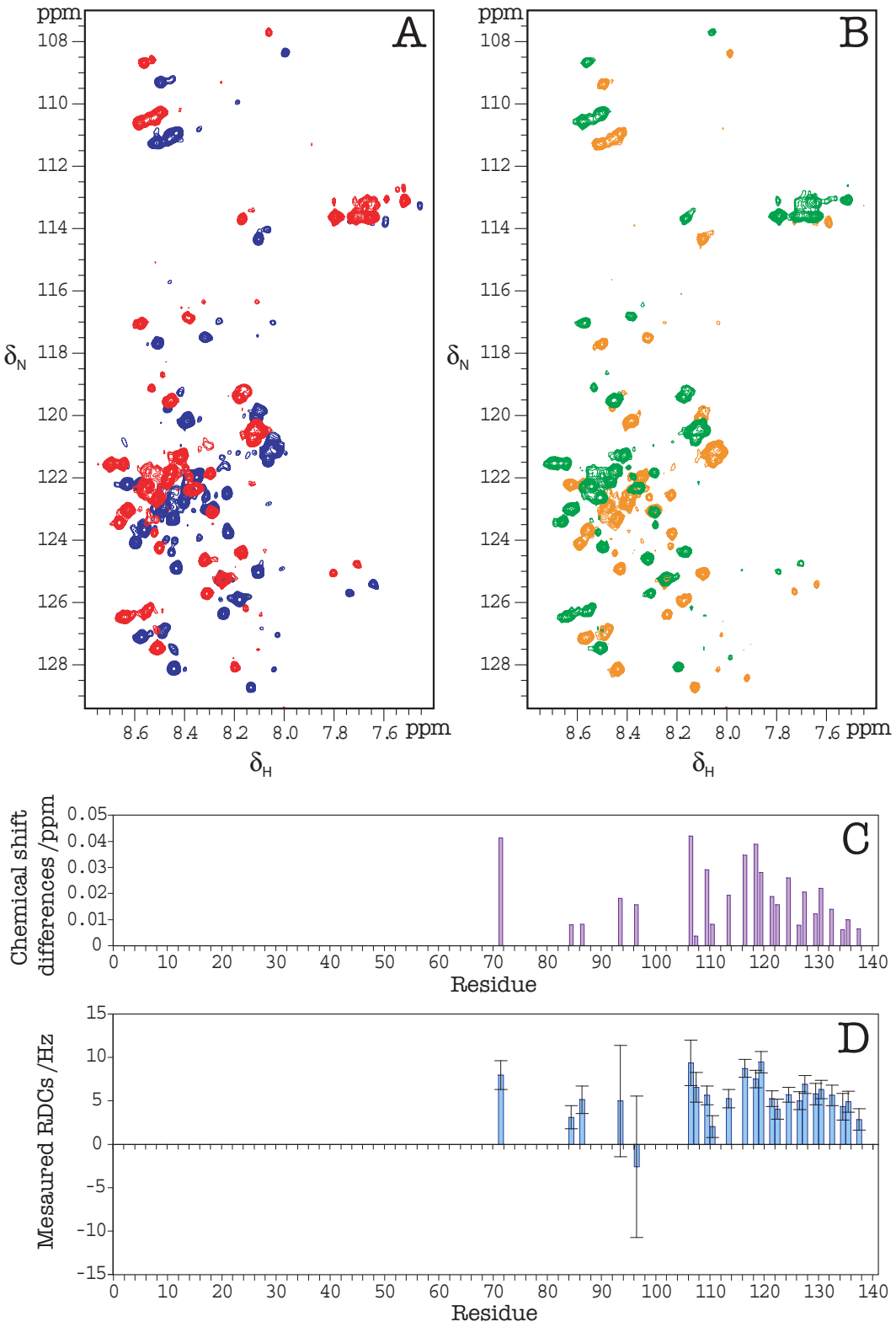

Figure 3.28: **A:** Isotropic ^{15}N HSQC (red) and TROSY (blue) spectra recorded on the V82M-αSyn RNC. **B:** ^{15}N HSQC (green) and TROSY (orange) spectra recorded on V82M-αSyn RNC aligned in phage. **C:** Plot of chemical shift changes for the V82M-RNCs upon addition of phage. **D:** Plot of measured N-H RDCs for the V82M-αSyn RNC.

3.2.3.5 Investigation of interactions between the ribosome and αSyn NCs

To examine the possibility that the broadening of the N-terminal signals was the result of interactions between the αSyn NC and the ribosome, further analysis was undertaken. In the first instance, a western blot against the SecM motif was used to further assess the integrity of the RNC both before and after NMR (Figure 3.29). This was performed to ensure that the NC was complete intact, and as one possibility for the absence of the N-terminal resonances would be that proteolysis has degraded the N-terminus of the NC. The anti-SecM western blot probes the SecM sequence present at the C-terminus of the αSyn sequence, which is protected within the ribosomal exit tunnel in RNCs. Therefore any proteolysis at the N-terminus of the NC would be manifested on the blot as bands of lower molecular weight than the full-length NC, and a reduction in the intensity of the full-length αSyn band.

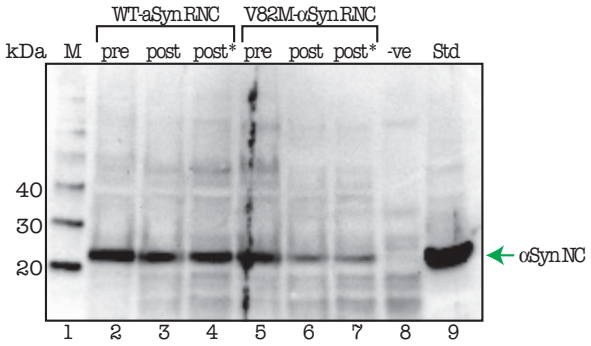

Figure 3.29: Anti-SecM western blot of samples from αSyn RNC samples under different conditions. M: marker, pre: sample before NMR acquisition, post: sample after NMR acquisition (140 hours for WT and 280 hours for V82M), post*: attached NC fraction of the sample post NMR acquisition, -ve: negative control (pellet of cushion loaded with purified, released V82M NC mixed with 70S ribosome), Std: standard of purified, released V82M NC. The bands corresponding to the NCs are indicated with the green arrow.

As seen previously in blots using the anti-α-Synuclein antibody (section 3.2.3.1), the αSyn NC migrates a little higher (above 20 kDa) than its true molecular weight of 18 kDa. The position of the αSyn NC band is confirmed with a standard of purified, released V82M-αSyn NC (lane 9, Std, Figure 3.29), which migrated at the expected position for the full-length NC (Figure 3.29, green arrow).

The same approach was applied for both WT- and V82M-RNCs. Initially, a sample was taken before NMR acquisition (‘pre’), representing the attached, intact NC. Another

sample was taken directly after NMR acquisition and loaded on to the blot without any further processing ('post') to represent the total NC content, including attached and any released NC or cleaved products. In addition, a sample taken after NMR acquisition was pelleted through a sucrose cushion, from which the pellet fraction ('post*') contains only the attached NC (Figure 3.29). Due to the observation of background binding of the antibody to non SecM-containing proteins, a negative control (-ve) was also used; this was prepared from the sucrose-cushion pellet of a sample in which isolated V82M was added to 70S ribosomes prior to the cushion. This negative control showed that no released NC could be carried over into the pellets and illustrated the pattern of weak bands resulting from non-specific bindings of the antibody to ribosomal protein for comparison with the RNC samples.

For the WT-αSyn RNC, all three samples gave NC bands of similar intensity, indicating that the NC remained bound to the ribosome throughout NMR data collection. In particular, the NC bands in the 'post' and 'post*' samples were of comparable intensity. This indicates that the NC population present in the sample after NMR ('post') is fully ribosome bound ('post*'). For the V82M-αSyn RNC, there is a small difference between the samples from before ('pre') and after ('post') NMR acquisition. However, the NC content ('post') of the sample after NMR is ribosome-bound ('post*'). In addition, none of the samples show evidence of proteolytic cleavage, suggesting that the RNCs were indeed intact and ribosome-bound. As the western blot indicated that the RNCs remain intact throughout NMR data acquisition, the origin of the line-broadening that leads to the loss of N-terminal peaks in the RNC HSQC spectra may be interaction between the NC and the ribosomal surface.

The flexibility of the αSyn protein is such that the NC could bend back on itself as a result of an interaction between the ribosome and the N-terminus of the NC (Figure 3.30). Such an interaction between the NC and the ribosomal surface could arise from electrostatic attraction. The first 100 residues of αSyn are positively charged [251] (due to the repeat KTKEGV motif), whereas the C-terminus is negatively charged and the NAC domain contains a hydrophobic stretch of amino acids. The surface of the ribosome is comprised of both RNA and proteins, with the RNA molecules being negatively charged giving an

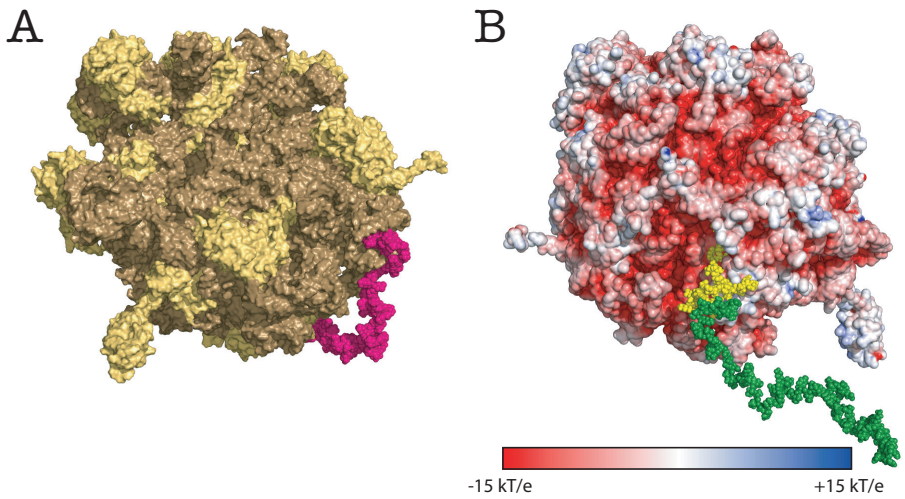

Figure 3.30: A: A model of αSyn in which the NC is interacting with the surface of the ribosome at its N-terminus. Only the 50S subunit is shown with proteins in yellow and RNA in cream. The α-synuclein NC is shown in magenta. **B:** Model of the 50S with αSyn NC present. The surface of the ribosome is highly negatively charged (red), while the N-terminal 100 residues of the αSyn-NC are positively charged (green, theoretical pI 9.4), and the region close to the C-terminus (residues 100-130) is negatively charged (yellow, theoretical pI 3.55). (Models built by Dr Chris Waudby and Luke Goodsell, UCL)

overall negative surface charge. Therefore, it is possible that the N-terminus is attracted to the negatively charged ribosome surface, while no such attraction would exist for the C-terminus. There may also be additional contributions to the electrostatic attraction from partial positive charges on the N-terminal his-tag. Indeed, HSQC spectra recorded on the isolated WT protein that does not have a his-tag at the N-terminus showed weaker interaction with the ribosome (as observed over residues 1-40) than did the isolated V82M protein with his-tag (Figure 3.31 A&B). Further tests for the presence of electrostatic interactions were undertaken in collaboration with Annika Weise, UCL.

To attenuate electrostatic interactions, the most commonly used approach is to increase the ionic strength of the buffer, within which the increased concentration of ions then screen the charge of the molecules more effectively. Within the confines of conditions suitable for maintaining the stability of the ribosome, two buffer modifications were tested: an increase in the concentration of NH_4Cl from 30 mM to 140 mM and substitution of the NH_4Cl with a combination of 100 mM arginine and 100 mM glutamate, which has been shown to disrupt protein-protein interactions without compromising the sensitivity of cryo-probes

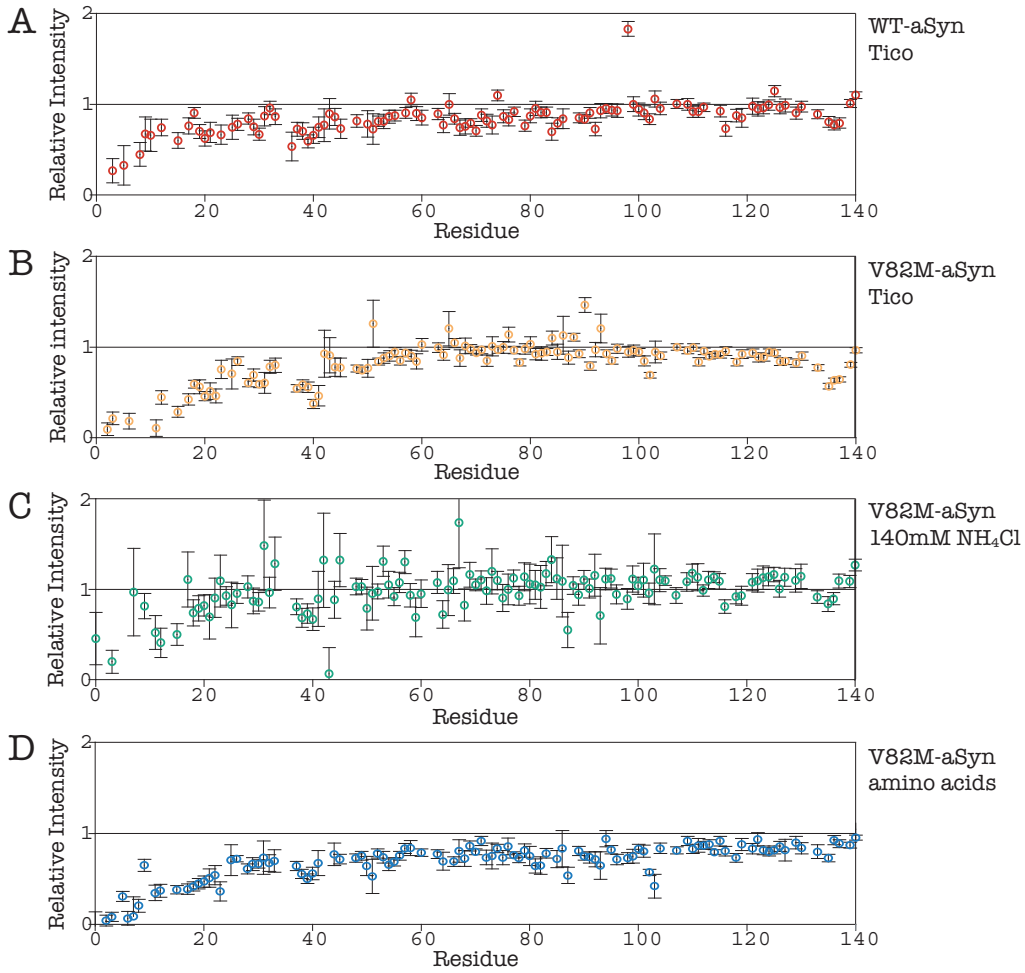

Figure 3.31: Plot of the relative peak intensities against residue number for αSyn in the presence of equimolar 70S ribosomes vs αSyn alone for (A) WT-αSyn (no his-tag) in Tico buffer, (B) V82M-αSyn in Tico buffer, (C) V82M-αSyn in Tico buffer with 140 mM NH_4Cl increased from 30 mM in Tico and (D) V82M-αSyn in Tico buffer with 100 mM arginine and 100 mM glutamate instead of NH_4Cl.

[252]. ^{15}N HSQC spectra and diffusion experiments were recorded on the ^{15}N-labelled isolated protein, unlabelled 70S ribosomes alone and the isolated protein in the presence of equimolar ribosomes (all at 5 μM concentration) using the buffer conditions described above. The original Tico buffer condition was also tested for comparison.

The data recorded using the 140 mM NH_4Cl buffer sample has lower signal-to-noise ratio, which is due to the reduced sensitivity of the cryo-probe in the presence of the higher concentration of ions. In Figure 3.31, the intensities of the signals from the isolated protein in the presence of ribosomes are plotted relative to the intensities from the isolated protein alone. For the WT-αSyn and V82M-αSyn isolated proteins in Tico buffer, it is clear that

the relative intensities are lower at the N-terminus than in the rest of the protein, which supports the hypothesis that there is an interaction between the N-terminus of the proteins and the ribosomal surface. If this interaction is assumed to be a fast-exchange event, the decrease in intensity of the N-terminal residues should correspond to an associated increase in the linewidth, which itself is a result of increase in the transverse relaxation rate R_2. This increase in R_2 has two possible contributions, the increase in the population-weighted average intrinsic relaxation rate (due to the much faster relaxation in the putative ribosome-bound form) and an additional exchange contribution, which arises from chemical shift changes between the free and bound forms. The population-weighted average intrinsic relaxation rate depends on the intrinsic relaxation rates of the free and ribosome-bound forms and also on the relative populations according to :

$$R^0_{2,\text{average}} = p_{\text{free}} R^0_{2,\text{free}} + p_{\text{bound}} R^0_{2,\text{bound}} \tag{3.1}$$

where $R^0_{2,\text{average}}$ is the average intrinsic relaxation rate, p_{free} and p_{bound} are the populations of the free and bound forms, and $R^0_{2,\text{free}}$ and $R^0_{2,\text{bound}}$ are the intrinsic transverse relaxation rates of its free and bound forms, respectively. Since $R^0_{2,\text{bound}} \gg R^0_{2,\text{free}}$, even a very low population of the bound state will give rise to appreciable line-broadening. This contribution to the line-broadening is likely to outweigh any exchange contributions that arise from possible chemical shift changes. In principle, it is also possible that the interaction falls into the slow exchange regime. In this case, only the free form of the protein will be observed, whose signals would then have relaxation properties similar to those of the isolated protein. However, the observed intensities indicate that if the interaction were to be in the slow exchange regime, the upper limit of the fraction bound would be 90 %, because the lowest observable intensity is 10 % of the free form. At the concentration used in these experiments (5 μM) and assuming a diffusion-limited on-rate, this corresponds to a K_d of 50 nM. Subsequent diffusion experiments (Figure 3.32), however, suggested a higher K_d of ~22 μM, so that the slow exchange regime scenario can be discounted.

The relative intensities of αSyn in 140 mM NH_4Cl buffer (Figure 3.31 C) show a similar

trend to those recorded in the original Tico buffer (Figure 3.31 B), although the recovery of the relative intensities by moving through the sequence from the N-terminus occurs within a shorter stretch of residues than for the sample in Tico buffer, indicating that the length of the interacting region is slightly shortened in the high-salt buffer. No such difference is observed in the arginine/glutamate buffer, for which the shape of the profile is indistinguishable from that in the original Tico buffer (Figure 3.31 D).

In order to quantify the strength of the interaction under the different conditions, diffusion experiments were recorded on the unlabelled 70S ribosome, and the isolated protein, both alone and in the presence of ribosomes. The diffusion coefficient measured for the isolated protein in the presence of ribosomes is a population-weighted average of the bound and free forms, and can be written:

$$D_{\text{obs}} = f_{\text{bound}} \cdot D_{\text{bound}} + f_{\text{free}} \cdot D_{\text{free}} \tag{3.2}$$

where D_{obs} is the population-weighted average diffusion coefficient, D_{bound} is the diffusion coefficient of the protein bound to the ribosome, D_{free} is the diffusion coefficient of the free protein, and f_{free} and f_{bound} are the relative populations of the free and bound forms, respectively. As the size of the ribosome is much greater than that of the protein, it can be assumed that $D_{\text{bound}} = D_{\text{ribosome}}$. In addition, $f_{\text{free}} = 1 - f_{\text{bound}}$. Since

$$K_{\text{d}} = \frac{[\text{protein}] \cdot [\text{ribosome}]}{[\text{protein-ribosome}]} \tag{3.3}$$

and since the total protein and ribosome concentrations are both 5 μM, it follows that:

$$K_{\text{d}} = 5 \times 10^{-6} \left(\frac{D_{\text{ribosome}} - D_{\text{obs}}}{D_{\text{obs}} - D_{\text{free}}} + \frac{D_{\text{obs}} - D_{\text{ribosome}}}{D_{\text{ribosome}} - D_{\text{free}}} \right) \tag{3.4}$$

The diffusion coefficient of the protein alone (D_{free}), the ribosome alone (D_{ribosome}) and the protein in the presence of ribosomes (D_{obs}) can be measured for all conditions, and hence the corresponding K_{d} values can be calculated according to Equation 3.4. The results are shown

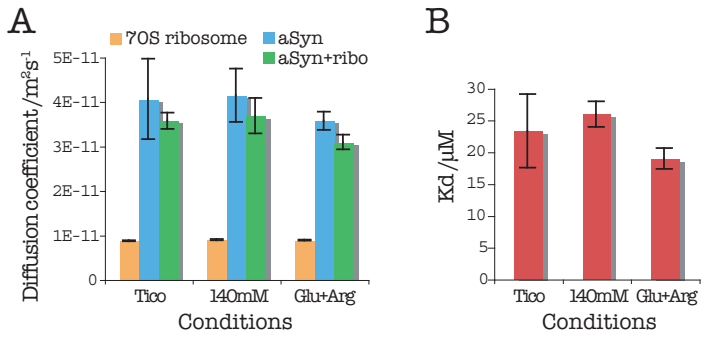

Figure 3.32: A: Plot of the diffusion coefficients for 70S ribosomes, αSyn protein, and αSyn protein in the presence of ribosomes under different buffer conditions. **B:** K_d of αSyn interacting with ribosomes derived from the diffusion constants.

in (Figure 3.32 B). The calculated K_d values are all higher than 20 µM, which implies k_{off} >1x10^5 Hz assuming that the on-rate is diffusion-limited. The associated k_{ex} is too large to be consistent with slow exchange on the chemical shift time-scale. Therefore, the interaction between the αSyn proteins and the ribosomal surface is a fast exchange event. The calculated K_d values are quite similar, although the interaction may be slightly weakened by the high-salt (140 mM NH_4Cl) buffer, in agreement with the small change observed in the relative intensity profile under these buffer conditions.

Nevertheless, it appears that even in the high-salt buffer, there remains some interaction between the N-terminus of αSyn and the ribosomal surface. This interaction could be further weakened by removal of the N-terminal his-tag. In the case that the residual interaction cannot be attenuated further, and still leads to loss of signals in the ^{15}N HSQC spectrum of αSyn, an alternative avenue of investigation could be to examine other intrinsically-disordered proteins that lack the repeated stretches of positively-charged residues found in αSyn.

3.3 Conclusions

This chapter has described the progress made towards measuring RDCs on ribosome bound-nascent chains, and has demonstrated the challenges that need to be overcome.

The development of the RNC expression and purification has been very successful, increasing the total yield of RNC from 100 pmol per litre to 3000 pmol per litre, an improvement which has enables the routine production of high-quality, high-concentration RNC samples, and has allowed investigation of RNCs using NMR methods, such as RDCs, possible. Taking advantage of these development, two types of RNCs were studied, the folded Dom5+110 RNC and the unfolded αSyn RNCs.

In the case of the Dom5+110 RNC, the sample produced was of high purity and relatively long lifetime, with 65 % of the NCs still attached after two days of NMR acquisition. From the analysis of linewidths and signal intensities, it was found that the motion of the folded Dom5 is significantly restricted by being anchored to the ribosome despite the length and expected flexibility of the linker. This restriction in domain motion gave rise to low signal intensities and broad linewidths, which precluded acquisition of high-quality RDC data.

However, the studies on the unfolded RNCs, based on αSyn, are more promising. αSyn RNCs are able to provide higher signal intensities and better sample stability, benefitting from recording the NMR experiments at a lower temperature. The challenges of the αSyn RNCs lie in the complexity of extracting structural information from the RDC data and the interaction between the NC and the ribosomal surface. Further work is required towards abrogating the interaction between the N-terminus of αSyn and the ribosome in order to recover the complete ^{15}N HSQC spectrum of the NC.

Chapter 4

Conclusions

4.1 Protein folding on the ribosome as studied by NMR spectroscopy

Much of our understanding of protein folding comes from *in vitro* studies of isolated proteins; within living systems, however, protein folding is known to be initiated during translation on the ribosome. Until very recently, studies of co-translational folding have been limited to biochemical and biophysical studies. Together, these have provided a wealth of information showing that nascent chains are capable of adopting native-like structure on the ribosome. Structural studies using X-ray crystallography and cryo-EM have provided detailed insights into the structure and function of the ribosome, but the highly dynamic nature of the emerging ribosome-bound nascent chain has made its characterisation impossible using these traditional strategies.

An understanding of the structural and dynamic transformations experienced by the nascent chain (NC) as it acquires its native structure would provide an unprecedented insight into the process of co-translational folding as it occurs on the ribosome. As discussed below, NMR spectroscopy offers the unique advantage of probing dynamic regions within the otherwise rigid ribosome. Specifically, NMR spectroscopy has allowed structural and dynamic characterisation of the highly mobile ribosomal L7/L12 stalk protein, which has

eluded both X-ray crystallography and cryo-EM [34]. These initial pioneering NMR studies have provided a solid foundation for the exploration of the emerging nascent polypeptide chain, because when bound to the ribosome, the nascent chain displays motional properties that are partly independent of the ribosome itself [124, 207, 253]. The inherent dynamic properties of the nascent chain therefore represent an attractive opportunity to study protein folding as it occurs on the ribosome, and provide a new avenue of investigation into protein folding as it takes place within living systems.

An idealised set of conditions for examining co-translational folding would involve the use of NMR spectroscopy to probe the emerging nascent chain during synthesis in 'real time'. As will be discussed later in this chapter, one significant limitation in the study of ribosome–nascent chain complexes (RNCs) by NMR spectroscopy is related to the low NMR sensitivity associated with these samples. Therefore, at present, co-translational folding studies using NMR are performed as equilibrium measurements; homogeneously stalled and isotopically labelled ribosome–nascent chain complexes permit examination of the nascent chain at a fixed length. By studying RNCs of varying lengths, folding 'snapshots' are generated which represent the real-time folding process. Research within the group has led to the development of strategies that enable RNCs to be produced *in vivo* within *E. coli* for NMR studies. Developing further methods by which direct structural evidence can be gleaned from NMR studies of RNCs is an important step towards understanding co-translational folding. In order to implement structural strategies that report on co-translational folding, this thesis has explored the development of residual dipolar couplings (RDC) applied to both the ribosome and RNCs to generate direct structural information.

This work builds upon recent NMR studies of ribosome–nascent chain complexes of an immunoglobulin domain, ddFLN. These studies have permitted initial chemical shift analyses, making steps towards probing structural and dynamic properties of the ribosome-bound nascent chain. These are beginning to provide a view of the nascent chain as it exists on the ribosome and the manner in which it acquires its structure during co-translational folding. At present, however, there is limited direct three-dimensional structural information

on the NC and as a consequence little is known about either the orientation of the NC with respect to the ribosome, or its structural preferences as it progressively emerges from the ribosome. Detailed studies reporting on these characteristics of the NC would provide a more comprehensive understanding of the early events of protein synthesis and folding.

4.2 RDCs applied to the study of the isolated and ribosome-bound L7/L12

Residual dipolar couplings are NMR parameters that can be measured with high sensitivity and that provide high-resolution structural information, in terms of both the orientations of individual inter-nuclear vectors within the protein structure and the global orientational probability distribution of the protein under anisotropic conditions. In order to explore the possibility of applying an RDC approach to the study of NCs, the technique was first applied to the dynamic L7/L12 stalk region of the ribosome, with a view to developing the required methodology for the subsequent study of RNCs. The rationale behind this strategy was based on the similar motional properties that are expected for the two dynamic, ribosome-bound species.

As described in Chapter 2, the development of an RDC strategy to derive alignment tensors and local structural information for L7/L12 on the 70S ribosome was largely dictated by the challenges associated with the low maximum working concentration of the ribosome (10 μM). This consideration narrowed the choice for the types of RDC that could be measured, with the focus on intrinsically large couplings (N-H and C^{α}-H^{α}) and those that, while smaller, can be measured with high sensitivity (N-C' and C'-C^{α}).

The alignment of the ribosome was successfully achieved using phage, and with the above considerations in mind, N-H RDC measurements were acquired using an HSQC/TROSY strategy. This approach was found to be optimal in this case because of the associated sensitivity and resolution gains compared to the IPAP-HSQC approach, which is more commonly applied to small, isolated proteins. The use of C^{α}-H^{α} couplings was

also explored because of their intrinsically large size. Modifications of established pulse sequences were implemented in an attempt to optimise the sensitivity and resolution for examining the ribosome-bound L7/L12.

To complement the N-H measurements made in phage, efforts to acquire RDCs in an alternative alignment medium were also made. Due to its large size and intrinsic lability, aligning the ribosome is not as straightforward as for simple proteins. The Otting medium, based on polyethylene glycol, was selected as a promising candidate for a second alignment medium, since it is reported to be inert towards most proteins, and should provide a near orthogonal alignment tensor to that produced in phage. However, the alignment was not as effective as with phage, as the ribosome appeared to interfere with the formation of the liquid crystal phase, although further attempts at lower concentrations could prove successful.

The N-H RDCs measured in phage using the HSQC/TROSY approach were able to yield detailed information on the ribosome-bound L7/L12. The measurement of the C^{α}-H^{α} RDCs was more challenging, with limited resolution in the $^{13}C^{\alpha}$-HSQC of ribosome-bound L7/L12 preventing extraction of splittings for many of the residues. The measurement of isotropic splittings for the H-C'/N-C' couplings showed that it may be feasible to measure the N-C' RDCs with sufficient precision and sensitivity to provide useful information. An attempt to measure the C'-C^{α} RDCs was unsuccessful, due to the low sensitivity associated with the ribosome-bound L7/L12. Indeed, the results of this study suggest that given the current sample conditions suitable for ribosomes, a future prospect for the measurement of ribosomal RDCs may lie in an alternative choice of organism. Thermophilic ribosomes (*e.g.* from *Thermus thermophilus*) have proved to be instrumental in crystallographic studies, presumably due to the inherent stability of the complexes. The enhanced thermostability may present the opportunity to either increase the concentration or, more likely, extend the lifetime of the 70S complex. As the L7/L12 stalk protein is highly conserved, presumably so too are its dynamic properties, and hence the development of RDCs in this context may offer alternative avenues of exploration.

The studies of the ribosome-bound L7/L12 were undertaken in parallel with isolated

L7/L12, allowing a comparison of its structural characteristics in isolation and when bound to the ribosome. A concentration-dependent oligomerisation of isolated L7/L12, postulated to be a dimer-tetramer interconversion, was observed. The oligomerisation event appeared to be mediated by the N-terminal domain of the L7/L12 protein and complicated RDC analysis due to the exchange between the different oligomeric states. For ribosome-bound L7/L12, the NTD is NMR-invisible and its structure could not be studied. The interpretation of the RDC data required careful consideration to take into account the intrinsic dynamics of L7/L12, so that only data from the rigid regions of the protein were considered.

The study of the CTD of isolated and ribosome-bound L7/L12 provided insights into both its structure and orientational preferences in the two states. The alignment tensors of the CTDs from the two gel-filtration fractions were very similar. This result is likely to arise from a combination of the weak alignment of the different NTDs and the fact that the alignment of the CTD will be dictated primarily from its own interaction with the phage, so that any influence from the NTD that may be transmitted through the linker region is so small as to be undetectable. The alignment tensors for the isolated and ribosome-bound CTDs were slightly different (although not to a highly significant degree), which may provide evidence for the orientation of the ribosome, and hence ribosome-bound NTD, affecting that of the L7/L12 CTD. Interestingly, however, only one set of RDCs was observed for the ribosome-bound L7/L12, despite the current understanding that the NMR signal arises from two CTDs, one on each of the two L7/L12 dimers. This suggests that there may be exchange of the dimers between the two ribosome-binding sites on the L10 protein or that the influence of the ribosome on the orientational preference of the extended CTDs is very small. Further studies using lanthanide-based alignment should yield more information on the flexibility of the linker region by revealing more clearly the relative orientational preferences of the N- and C-terminal domains.

The measurement of the N-H RDCs in phage also indicated that the structure of the L7/L12 CTD is very similar in the isolated and ribosome-bound states, and this similarity was confirmed by a limited structural refinement using the RDC data. Improvement of the

refinement protocol requires measurement of the N-H RDCs in another alignment medium. Attempts in this direction with the Otting medium were as yet only partly successful, but with indications that modifications could lead to improved alignment.

4.3 RDCs applied to study of RNCs

As mentioned, the primary purpose of studying the ribosome-bound L7/L12 was to provide a foundation of optimised methodology that could then be applied to the ribosome-bound nascent chain. An inherent challenge in the study of RNCs, however, is the preparation of suitable samples. During the period of this study, significant improvements were made that enabled the generation of large quantities of isotopically labelled and homogeneously stalled RNCs, with an enhanced stability on the order of days. The use of the model RNC system Dom5+110 was initially selected as a candidate for RDC experiments because of its favourable characteristics as an RNC, such as its high occupancy and stability. The dynamic properties of the RNC, however, yielded resonances with lower than expected intensities, which presented a significant challenge in obtaining the required sensitivity and resolution to measure peak positions with the precision required for N-H based RDC experiments.

As an alternative, the RDC study of αSyn-RNCs was initiated because the intrinsically disordered conformation of αSyn and its high degree of flexibility gives rise to intense and narrow resonances. In addition, the low temperatures used for such experiments extended the sample lifetime, allowing for longer NMR data acquisitions. N-H RDC experiments were recorded on the V82M αSyn-RNC, giving rise to a collection of RDCs corresponding to residues at the C-terminus but little data from residues in the N-terminal half of the protein, whose signals are broadened beyond detection. This broadening appears to arise as a result of an interaction between the N-terminus of αSyn and the ribosome. The nature of the interaction is likely to be electrostatic in origin due to the net positive charge associated with the N-terminus of αSyn, and it was found that the interaction could be weakened slightly by high ionic-strength conditions. Further studies into the nature of the interaction and the means by which it can be attenuated or abrogated may permit a more complete set of

RDC data to be obtained. Alternatively, other intrinsically disordered proteins could be investigated to determine if they can be used as model systems for intrinsically disordered RNCs.

4.4 Future strategies for the application of RDCs to RNCs

As detailed above, the relatively narrow linewidths and high signal-to-noise ratios associated with the inherent dynamics of intrinsically disordered RNCs make them more amenable systems for developing RDCs at present. However, more attractive proposition for developing our understanding of co-translational folding is the study of folding-competent proteins. Returning to the case of the Dom5+110-RNC, the application of RDCs may be more feasible if an alternative isotopic labelling strategy is considered. Specifically, the use of perdeuteration and selective methyl labelling of isoleucine, leucine and valine amino acids enable methyl-TROSY methodology to be applied, which both enhances the sensitivity of the resonances and reduces their associated linewidths. This approach has been used recently to acquire spectra of selectively isoleucine-labelled Dom5-RNCs (Dr Helene Launay, UCL). It also proved to be critical in overcoming the motional restrictions imposed on the NCs of Dom5-RNCs that were close to the ribosomal exit tunnel [207], and enabled identification of native-like cross-peaks of folded Dom5 that were otherwise not observable using uniform ^{13}C or ^{15}N strategies. Developing RDCs along similar lines may therefore present new opportunities.

It must be noted, however, that methyl groups at the termini of amino-acid side-chains show a wide range of local flexibility relative to the backbone and hence application of methyl RDCs to structural refinement also requires careful characterisation of the local dynamics by relaxation measurements. Despite this, the use of methyl RDCs has been explored recently in a small number of studies, for example, on ubiquitin [254] and malate synthase G [255], the latter an 82-kDa protein. With this in mind, the use of methyl RDCs for a selectively-isoleucine labelled Dom5+110-RNC presents an elegant case for the application of this strategy to derive information on co-translational folding.

4.5 Prospects for NMR and RDCs in the study of co-translational protein folding

The use of RDCs to provide direct structural information will undoubtedly complement the existing strategies, in which the study of NCs by NMR proceeds principally via monitoring the chemical shifts and comparing them with those of the isolated protein at different folding stages between fully folded and fully denatured conformations. Although the backbone amide chemical shifts have, up to now, been the main source of chemical shift information on NCs, a more recent study (Dr Helene Launay, UCL) has shown that methyl TROSY of selectively labelled NCs can also provide improved signal-to-noise ratios.

Although chemical shift analyses are able to provide useful indicators of the folding status of the NC, the information content of chemical shifts is currently somewhat limited as the conversion of chemical shift information into structural restraints is not yet straightforward. In this respect, RDCs are a more powerful tool, as they are directly translatable into structural restraints. Although RDCs are not the only NMR restraints that can be used in structure calculations, they can be measured using relatively sensitive NMR methods, an important consideration for application to RNCs.

Together, the combined approach of direct structural information (from RDCs) with indirect chemical shift information will provide a powerful strategy for the future study of RNCs, for example by applying these data within NMR restraint-driven molecular dynamics simulations to generate ensemble structures. Indeed, the use of chemical shifts alone is currently providing a means of generating such ensembles and is already being applied to RNCs of ddFLN. The further development and use of RNC-derived RDCs in combination with chemical shift information offers the exciting possibility of providing further insights into the emerging nascent chain and the development of its structure at the very early stages of folding. RDCs offer the possibility of reporting on interactions the NC may make with the ribosome as well as the prospect of examining NC-protein interactions, such as those made with auxiliary factors such as molecular chaperones (trigger factor), those involved

in NC transport (*e.g.* SecA/signal recognition particle) or small molecules or peptides that may stabilise a preference for a particular conformational state of the NC. Overall, probing the structure and orientation of emerging NCs by RDCs will provide a wealth of highly complementary structural information that will make important contributions to our understanding of protein folding as it takes place on the ribosome.

Chapter 5

Materials and Methods

The following chapter describes the methodology used for the studies discussed in this thesis. It contains the protocols for the preparation of *E. coli* ribosomes (both unlabelled and isotopically labelled) and ribosome-nascent chain complexes (RNCs) that were used in the NMR studies, the biochemical methods used to assess the purity and integrity of the ribosomes and RNCs and the NMR experiments and data analysis associated. All biochemicals and reagents were purchased from sigma, unless otherwise specified and were of the highest grade possible.

5.1 Molecular biology and biochemical methods

5.1.1 Vectors and sequences

All constructs (except for L7/L12 and αSyn) used were cloned into the pLDC vector [253], which contains a T7 promoter, N-terminal his tag and ampicillin resistance. The $ddFLN_{646\text{-}838}$ RNC and $ddFLN_{646\text{-}838}$ isolated protein constructs were constructed by Dr Lisa Cabrita. The $ddFLN_{646\text{-}750}$ protein construct was made by site-directed mutagenesis of the $ddFLN_{646\text{-}838}$ construct. The $ddFLN_{646\text{-}750}$-LBT and $ddFLN_{646\text{-}838}$-LBT RNC constructs were cloned by insertion of the LBT DNA sequence into the existing construct between the *Nde*I and *Nhe*I restriction sites by cassette mutagenesis, in which a pair of complementary

oligonucleotides were annealed and ligated into the pLDC vector. The clone of the L7/L12 protein (ASKA clone JW3949) was obtained from the National BioResource Project (NIG, Japan) and then cloned into the pET21b(+) vector with a C-terminal his-tag between the *Nde*I and *Xho*I restriction sites (clone produced by Dr Lisa Cabrita). The clone of isolated αSyn protein was in the pT7-7 vector and a kind gift from Peter Lansbury (Harvard University).

5.1.2 Bacterial strains

The bacterial strain DH5α (Invitrogen) was used for cloning. For protein, ribosome and RNC expressions, bacterial strain BL21-GOLD(DE3) (Stratagene) was used. Both bacterial strains were purchased as competent cells.

5.1.3 Growth media

Lysogeny broth (LB) medium (Table 5.1) was used for routine bacterial growths including overnight cultures, transformations, agar plates, unlabelled protein, ribosome and RNCs expression. For making agar plates, the medium was supplemented with 15 g/l of agar. The medium was sterilised by autoclaving (121 °C, 20 min) prior to addition of antibiotics. All constructs discussed in this work are ampicillin resistance: 100 μg/ml ampicillin was added to the medium after sterilisation when the medium has cooled to below 50 °C, except for ribosome production, in which no antibiotic was required.

Table 5.1: LB medium composition

Compound	Concentration
Yeast extract	5 g/l
Tryptone	10 g/l
NaCl	10 g/l

Minimal medium (M9) (Table 5.2) was used for expression of isotopically labelled proteins. M9 salts were made as 10× stock at pH 7.4, and all components were added to autoclaved distilled water under sterile conditions.

Compound	Concentration	sterilisation
M9 salts	1 X	autoclave
$MgSO_4$	2 mM	filter
$CaCl_2$	100 mM	filter
yeast nitrogen base†	0.8 g/l	filter
$^{15}NH_4Cl$	1 g/l	filter
glucose*	0.4 % (w/v)	filter

Table 5.2: M9 medium composition. *0.2% ^{13}C glucose was used if ^{13}C labelling was required. †Without amino acids, Difco.

Compound	Concentration
Na_2HPO_4	71 g/l
KH_2PO_4	34 g/l
NaCl	58.4 g/l

Table 5.3: 10X M9 salts composition, pH 7.4, autoclaved for sterilisation.

Ribosome and RNC were produced in *E. coli* using a two stage growth and expression protocol. MDG medium (Table 5.4) was used in both ribosome and RNC production in the first stage to achieve high cell density. For expression of isotopically labelled ribosomes, L-aspartic acid was removed from the MDG recipe to achieve 100% labelling and $^{15}NH_4Cl$ was added separately to the medium being part of the 50× salt solution.

Compound	Concentration	sterilization
50X MDG salts	1 X	autoclave
5% (w/v) L-aspartic acid pH 7.0	0.2 %(w/v)	
$MgSO_4$	2 mM	filter
Trace metals*	0.02 % (w/v)	filter
glucose	0.4 % (w/v)	filter

Table 5.4: MDG medium composition. *100% trace metals: 2 mM $CoCl_2$, 2 mM $CuCl_2$, 2 mM $NiCl_2$, 2 mM Na_2SeO_3, 2 mM Na_2MoO_4, 2 mM H_3BO_3, 10 mM $ZnSO_4$, 10 mM $MnCl_2$, 20 mM $CaCl_2$, 50 mM $FeCl_3$.

Compound	Concentration
Na_2HPO_4	1.25 M
KH_2PO_4	1.25 M
Na_2SO_4	250 mM
NH_4Cl	2.5 M

Table 5.5: 50X MDG salts composition. Autoclaved for sterilisation.

Enhanced M9 (EM9) medium was used for the expression stage of RNC production. The composition of EM9 medium (based on [256]) is very similar to that of M9 medium, with the following modifications: the pH of the salts is 8.0-8.2, the concentration of $MgSO_4$ is 5 mM instead of 2 mM, 0.0125 % (v/v) of trace metals is added, BME vitamins (Sigma) is used instead of yeast nitrogen base and the concentration of $CaCl_2$ is 0.2 mM instead of 100 mM.

5.1.4 Transformation

To 25 µl of competent *E. coli* cells, 1 µl of a plasmid was added. The cells were incubated on ice for 30 min, followed by a heat-shock at 42 °C for 45 s. The cells were then incubated on ice for 2 min before adding 900 µl of LB media and incubating for 1 hour at 37 °C, 200 rpm shaking to allow for expression of the antibiotic resistance gene. The cells were then plated on LB-Amp plates and incubated at 37 °C for 12-16 hours, until the colonies were 2-3 mm in size.

5.1.5 Mutagenesis

QuikChange site-directed mutagenesis was used to modify the Dom5+110-RNC construct to generate the Dom5 construct. Cassette mutagenesis was used to add the lanthanide-binding tag to the Dom5 construct.

Oligonucleotide sequences For the addition of the lanthanide binding tag to Dom5 using cassette mutagenesis, complementary oligonucleotides were purchased from Eurofins (Germany) (high-purity, salt-free grade). The pair of oligonucleotides used for the mutagenesis had the LBT flanked by restriction enzyme sites *Nde*I (5') and *Nhe*I (3') as described by the following sequences. Each sequence is written in the 5' to 3' direction:

5’-3’: TATG CAT CAC CAT CAC CAT CAT TAT ATT GAT ACC AAT AAT GAT GGC TGG TAT GAA GGC GAT GAA CTG CTG GCG G

5’-3’: CTAGC CGC CAG CAG TTC ATC GCC TTC ATA CCA GCC ATC ATT ATT GGT ATC AAT ATA ATG ATG GTG ATG GTG ATG CA

For the introduction of the stop codon to convert Dom5+110-RNC construct to the Dom5 construct the following complementary sequences were used:

5’-3’: GTA CTT GGT AAC TAA AGT GAA TTC TTC

5’-3’: GAA GAA TTC ACT TTA GTT ACC AAG TAC

5.1.6 Polymerase chain reaction (PCR)

Amplification of DNA was performed by the polymerase chain reaction (PCR), in a Corbin Research Thermal Cycler, using the oligonucleotides presented above.

Table 5.6: PCR reaction components for KOD polymerase

Reagents	Stock	Final concentration	Volume
DNA template	40 ng/µl	10-500 ng	1 µl
5’ primer	10 µM	300 nM	0.75 µl
3’ primer	10 µM	300 nM	0.75 µl
dNTPs	2 mM	400 µM	5 µl
Polymerase	3 u/µl	3 u	1 µl
Polymerase buffer	2×	1×	12.5µl
ddH_2O	—	—	4 µl

Table 5.7: PCR reaction temperature cycling

Step	Temperature	Time	Cycles
Initial denaturation	95 °C	2 min	1
Denaturation	98 °C	10 s	
Annealing	50 °C	30 s	35
Extension	68 °C	7 min	
Final extension	72 °C	10 min	1
Hold	4 °C	—	—

The PCR products were verified by 0.8 % (w/v) agarose gel electrophoresis before proceeding further.

5.1.7 DNA purification and quantification

Cultures from 5 ml of overnight growth in LB were used to purify DNA using the QiaPrep Spin miniprep kit (Qiagen) following the manufacturer's instructions. For DNA quantification, the absorbance of the DNA was measured at 260 nm, and the concentration calculated using the following relation: $1OD_{260} \equiv 50$ ng/µl.

5.1.8 Restriction enzyme digestion and ligations

5.1.8.1 Restriction enzyme digestion

Restriction enzyme digestion of Dom5 (in the pLDC vector) were performed using enzymes and buffers from New England Biolabs as described below:

Table 5.8: *Nhe*I and *Nde*I digestion reaction components

Reagent	Volume /µL
DNA (400 ng/µl	30
*Nhe*I (NEB)	2.5
*Nde*I (NEB)	2.5
BSA (100×)	0.5
Buffer 4 (10×)	5
ddH_2O	9.5

For *Dpn*I digestion, 20 µl of PCR products were digested with *Dpn*I in a final reaction volume of 25 µl.

Digestions were typically incubated for 3 h at 37 °C before heat inactivation at 65 °C for 20 min.

5.1.8.2 Oligonucleotide annealing for cassette mutagenesis

Oligonucleotide pairs corresponding to the lanthanide binding tag were annealed to generate a cassette for ligation into the pLDC vector backbone. Equal proportions of forward and reverse oligonucleotides at a final concentration of 10 µM were prepared in ddH_2O. The oligonucleotides were then heated to 95 °C for 5 min and then allowed to cool slowly to room temperature prior to ligation.

5.1.8.3 Ligations

T4 DNA ligase and 10×T4 DNA ligase buffer (New England Biolabs) were used for the ligation reactions. Molar ratios of 1:3, 1:5, 1:10 vector-to-annealed oligonucleotides were used for cohesive end ligations. The reactions were incubate at 16 °C for 14-16 h, followed by heat deactivation of the T4 DNA ligase at 70 °C for 20 min.

5.1.9 Agarose gel electrophoresis

Reagents:

- 50× TAE buffer (for 1 l): 242 g Tris, 57.1 ml glacial acetic acid, 100 ml 0.5 M EDTA at pH 8.0.
- 6× loading dye (Promega)
- 1 kB ladder (Promega)
- Ethidium bromide 10 mg/ml

To 1× TAE, agarose was added to a final concentration of 1 % (w/v) and dissolved by heating. The solution was cooled and ethidium bromide was added to a final concentration of 1 µg/ml. The solution was poured into the gel frame with a comb and was allowed to set. The gel frame was then placed into the tank containing 1× TAE. The samples were mixed with an equal volume of loading dye and loaded into the wells. Electrophoresis was performed at a constant voltage of 100 V until the dye front had reached the bottom of the gel. The gel was then visualised under ultraviolet light.

Gel purification After electrophoresis, the DNA fragment of interest was excised from the gel and purified using the QiaQuick Gel Extraction kit (Qiagen) according to the manufacturer's instructions

5.1.10 Protein expression

All isolated proteins were expressed following the method below unless specified otherwise. A single colony was used to inoculate 3 ml of LB medium. The pre-culture was incubated overnight at 37 °C, with 200 rpm shaking. 500 ml of M9 medium prepared with $^{15}NH_4Cl$ and/or ^{13}C-glucose was inoculated with the pre-culture and incubated for 6 hours at 37 °C, with 200 rpm shaking. When the OD_{600} reached 0.6, IPTG was added to the medium to a final concentration of 1 mM to induce protein expression. After 4 h incubation at 37 °C (or 16-18 h at 21 °C), the cells were harvested by centrifugation at 6000 rpm for 20 min and the cell pellet transferred to a 50 ml tube for storage. Cells were flash-frozen in liquid nitrogen and stored at -20 °C until purification.

5.1.11 Protein purification

5.1.11.1 Purification of isolated ddFLN$_{646\text{-}750}$ and ddFLN$_{646\text{-}750}$-LBT

Buffer compositions

- Phosphate buffer: 25 mM NaH_2PO_4, 30 mM NH_4Cl, 2 mM BME, pH 7.4
- Phosphate washing buffer: 25 mM Iimidazole, 25 mM NaH_2PO_4, 30 mM NH_4Cl, 2 mM BME, pH 7.4
- Phosphate elution buffer: 150 mM imidazole, 25 mM NaH_2PO_4, 30 mM NH_4Cl, 2 mM BME, pH 7.4

Purification procedure The cells were resuspended in phosphate buffer supplemented with trace DNAase and 1 protease inhibitor tablet (sigma), and lysis was performed using sonication, performed on ice with interleaved periods of 10 s of sonication and 30 s of rest (10 cycles). The cellular debris was then pelleted by centrifugation at 18000 rpm for 45 min. The cell lysate was added to 5 ml Ni-NTA beads, and binding allowed to proceed for 1 hour at 4 °C. The beads were washed with phosphate washing buffer until baseline (measured using Bradford reagent: 160 µl reagent + 40 µl sample in a microplate), after which the protein was

eluted with phosphate elution buffer. The sample was then buffer exchanged into phosphate buffer and loaded onto a 5 ml Hitrap Q-sepharose column (GE Healthcare) using the Akta system (GE Healthcare). The Q-sepharose column was washed with phosphate buffer until baseline and the protein was eluted with a 60 ml salt gradient from 0-1 M NaCl at 5 ml/min flow-rate with collection of 1.5 ml fractions. The sample was then concentrated with a 10 kDa molecular-weight cut-off concentrator and run through a Superdex 75 16/60 column (GE healthcare) in phosphate buffer at 1 ml/min flow-rate with collection of 1 ml fractions. Pure fractions of Dom5 (as assessed by SDS-PAGE) were pooled and concentrated using a 10 kDa molecular-weight cut-off concentrator. The final concentration of protein was determined from the absorbance at 280 nm, using the extinction coefficient for Dom5 (ϵ=5960 $M^{-1}cm^{-1}$, calculated from amino-acid sequence). Released NCs of Dom5+110 extracted from the purification of RNCs were purified using the same protocol for the Ni-NTA chromatography.

5.1.11.2 Purification of isolated L7/L12

Buffer compositions

- Lysis buffer (50 ml): 50 mM Na_2HPO_4, 100 mM KCl, 1 protease inhibitor tablet (Roche), trace DNAseI, pH 7.0
- Ni binding buffer: 50 mM Na_2HPO_4, 100 mM KCl, pH 7.0
- Ni wash buffer: 50 mM Na_2HPO_4, 100 mM KCl, 25 mM imidazole, pH 7.0
- Ni elution buffer: 50 mM Na_2HPO_4, 100 mM KCl, 250 mM imidazole, pH 7.0
- Ion-exchange buffer A: 50 mM Tris·HCl, pH 8.0
- Ion-exchange buffer B: 50 mM Tris·HCl, 1 M NaCl, pH 8.0
- Size-exclusion buffer: same as Ni binding buffer

Purification procedure The cells were defrosted in a 37 °C water-bath and resuspended in lysis buffer. Lysis was performed using sonication, performed on ice with interleaved

periods of 10 s of sonication and 30 s of rest (10 cycles). The cellular debris was then pelleted by centrifugation at 18000 rpm for 45 min. The cell lysate was added to 5 ml Ni-IDA beads pre-equilibrated in Ni binding buffer, and binding allowed to proceed for 1 hour at 4 °C. The beads were washed first with Ni binding buffer until baseline (measured using Bradford reagent, 160 µl reagent + 40 µl sample) and second with Ni wash buffer (also until baseline) after which the protein was eluted with Ni elution buffer. The eluate was then dialysed into ion-exchange buffer A overnight at 4 °C and loaded on a 5 ml Hitrap Q-sepharose column (GE healthcare) using the Akta system at 5 ml/min flow-rate. The Q-sepharose column was washed first with ion-exchange buffer A and second with 90 %:10 % buffer A:buffer B until baseline, after which the protein was eluted with a 100 ml salt gradient from 10-100 % ion-exchange buffer B with collection of 1.5 ml fractions. The absorbance at 230 nm was monitored. Fractions were examined by SDS-PAGE. Pure fractions were pooled and concentrated using a 3 kDa molecular weight cut-off concentrator and then ran through a Superdex 200 16/60 column at 1 ml/min flow-rate (GE healthcare) pre-equilibrated in size-exclusion buffer with collection of 1 ml fractions. Pure fractions of L7/L12 (as assessed by SDS-PAGE) were pooled and concentrated using a 3 kDa molecular-weight cut-off concentrator. The final concentration of protein was determined using the bicinchoninic acid (BCA) protein assay reagent (Pierce Biotechnology, Inc.) because the absence of aromatic residues in the sequence, results in a very weak extinction coefficient at 280 nm.

5.1.11.3 Purification of isolated α-synuclein

Buffer compositions

- Lysis buffer (50 ml): 100 mM Tris·HCl, 10 mM EDTA, 2 mM BME, 1 protease inhibitor tablet (Sigma), pH 8.0
- Ion-exchange buffer A: 25 mM Tris·HCl, 2 mM BME, pH 7.7
- Ion-exchange buffer B: 25 mM Tris·HCl, 2 mM BME, 1 M NaCl, pH 7.7
- Size-exclusion buffer: same as ion-exchange buffer A

Purification procedure The cells were defrosted in a 37 °C water bath and resuspended in lysis buffer. Lysis was performed first with two freeze-thaw cycles (flash-freeze in liquid nitrogen followed by defrosting in a 37 °C water bath) and then by sonication, performed on ice with interleaved periods of 30 s of sonication and 30 s of rest (6 cycles). The cellular debris was then pelleted by centrifugation at 18000 rpm for 45 min. The cell lysate was then transferred to a beaker and boiled for 20 min with stirring. The aggregated *E. coli* proteins were separated by centrifugation at 13500 rpm for 20 min at 4 °C. Streptomycin sulfate was added to the supernatant at 10 mg/ml concentration and stirred for 20 min at 4 °C. The solution was then centrifuged at 13500 rpm for 20 min at 4 textdegree C and the supernatant transferred to a beaker. Ammonium sulfate was added at a concentration of 400 mg/ml and stirred for 30 min at 4 °C. The solution was again centrifuged at 13500 rpm for 20 min at 4 °C, and the supernatant discarded. The pellet was resuspended in a minimal volume of ion-exchange buffer A and then dialysed against 4 l of water overnight at 4 °C. The dialysed protein solution was filtered and loaded on a 5 ml Hitrap Q-sepharose column (pre-equilibrated in Ion-exchange buffer A) (GE healthcare) using the Akta system at 5 ml/min flow-rate. The Q-sepharose column was washed with ion-exchange buffer A until baseline, after which the protein was eluted with a 60 ml salt gradient from 0-600 mM NaCl using ion-exchange buffer B with collection of 2 ml fractions. Fractions were examined by SDS-PAGE and pure fractions dialysed into water before concentration with a 3 kDa molecular-weight cut-off concentrator. The final concentration of protein was determined from the absorbance at 280 nm, using the extinction coefficient for Dom5 (ϵ=5960 $M^{-1}cm^{-1}$, calculated from amino-acid sequence).

5.1.12 Expression and purification of *Escherichia coli* 70S ribosomes

The protocol described here is the basic method for ribosome purification; modifications made for particular cases are specified in the relevant sections.

Ribosomes were prepared by growing *E. coli* cells in MDG medium containing $^{15}NH_4Cl$ as the nitrogen source for 20 hours at 37 °C. Cells were then centrifuged at 4000 rpm for

20 minutes at 4 °C. The cell pellet was gently resuspended in the same volume of fresh MDG medium and incubated at 37 °C. The cell density was checked every 15 minutes until a steady increase of OD_{600} was observed. Cells were then harvested by centrifugation at 6000 rpm for 15 minutes at 4 °C.

The resulting cell pellet was resuspended in 50 ml lysis buffer (30 mM HEPES, 12 mM $Mg(OAc)_2$, 1 M KOAc, 5 mM EDTA, 2 mM BME, trace DNAseI, pH 7.6) and lysed with a French Press (1000 psi, 3 passes). The cell lysate was centrifuged at 18000 rpm for 45 minutes at 4 °C. The supernatant was collected and loaded onto a 30 %(w/v) sucrose cushion prepared in a high-salt buffer (30 mM HEPES, 12 mM $Mg(OAc)_2$, 1 M KOAc, 5 mM EDTA, 5 mM ATP, one protease inhibitor tablet (Roche), 2 mM BME, pH 7.6). The cushion was then centrifuged at 40000 rpm for 15 hours at 4 °C in a Beckman Type 45Ti rotor. The resulting ribosomal pellet was resuspended in the same buffer as used for the cushion but without ATP and further purified through a linear density gradient of 10 - 35 % (w/v) sucrose in a Beckman SW28 rotor (22000 rpm, 16 hours, 4 °C). The gradient after centrifugation was then fractionated using a sucrose gradient machine (Teledyne Isco Foxy Jr.). Fractions containing 70S ribosomes were identified by SDS-PAGE. All fractions containing pure 70S ribosomes were concentrated and buffer exchanged into Tico buffer (10 mM HEPES, 12 mM $MgCl_2$, 30 mM NH_4Cl, 2 mM BME, pH 7.6).

The final concentration of protein was determined from the absorbance at 260 nm, using the extinction coefficient ϵ=4.2×10^7 $M^{-1}cm^{-1}$ (A_{260}=1 corresponds to a 70S ribosome concentration of 24 pmol/ml).

5.1.13 Expression and purification of ribosome-nascent chain (RNC) complexes

RNCs were prepared by growing *E. coli* cells (BL21 DE3) in unlabelled MDG medium (with 100 µg/ml ampicillin) for 16 hours at 30 °C. Cells were then harvested by centrifugation at 3000 rpm for 15 mins at 4 °C. The cell pellet was washed twice by resuspension in the same volume of 1x EM9 salts (25 mM Na_2HPO_4, 25 mM KH_2PO_4, 5 mM NaCl, pH 8.0) and

centrifugation at 3000 rpm for 15 mins at 4 °C. The washed cell pellet was then resuspended in the same volume of EM9 media with 100 µg/ml ampicillin containing 1 g/l $^{15}NH_4Cl$ as the nitrogen source and 1 mM IPTG. The cells were allowed to grow for 10 mins at 30 °C prior to addition of 150 mg/l rifampicin. The cells were then allowed to grow for 60 mins at 30 °C before harvesting by centrifugation at 4000 rpm for 20 mins at 4 °C.

The final cell pellet was resuspended in lysis buffer (50 mM HEPES, 12 mM $Mg(OAc)_2$, 500 mM KOAc, 2 mM BME, 100 mg/ml lysozyme, pH 7.6) and lysed by French Press (1000 psi, 3 passes). The cell lysate was centrifuged at 18000 rpm for 45 mins at 4 °C. The supernatant was collected and loaded onto a 30 % (w/v) sucrose cushion prepared in 'Buffer B' (unit[50]mM HEPES, 12 mM $Mg(OAc)_2$, 500 mM KOAc, 2 mM BME, pH 7.6) with 5 mM ATP. The cushion was then centrifuged at 42000 rpm for 3 hours at 4 °C in a Beckman Type 45Ti rotor. The resulting ribosomal pellet was resuspended in Buffer B with 5 mM ATP and then loaded onto 5 ml Ni-IDA resin that has been pre-equilibrated in Buffer B and binding allowed to proceed for 1 h 4 °C. The resin was washed with Buffer B until the A_{260} reading reached zero and then eluted with 150mM imidazole in Tico buffer. The eluate was concentrated using a 100 kDa molecular-weight cut-off concentrator and further purified through a linear density gradient of 10-35% sucrose in a Beckman SW28 rotor (22000 rpm, 16 h, 4 °C), 1000 pmol of ribosomal material was loaded on each gradient. The gradient after centrifugation was then fractionated using a sucrose gradient machine. Fractions containing the pure ribosomal matrial were identified by SDS-PAGE. All fractions of pure ribosomal material were concentrated and buffer exchanged into Tico buffer (10 mM HEPES, 12 mM $MgCl_2$, 30 mM NH_4Cl, 2 mM BME, pH 7.6).

The final concentration of the RNC was determined using the same method as for the ribosomes. The occupancy was determined as described in section 5.1.16.3.

5.1.14 Small-scale RNC expression and purification

Small-scale RNCs expression RNC samples for small-scale purification were prepared by growing *E. coli* cells (BL21) in unlabelled MDG medium with 100 µg/ml ampicillin for

16 h at 30 °C. Cells were harvested at 3000rpm for 15 min at 4 °C. The cell pellet was then washed twice by resuspension in the same volume of 1× EM9 salts and centrifugation at 3000rpm for 15 min at 4 °C. The washed cell pellet was resuspended in the same volume of EM9 medium and expression induced with 1 mM IPTG. The cells were allowed to grow for 10 min at 30 °C prior to addition of 150 mg/l rifampicin. The cells were then allowed to grow for 30-60 mins at 30 °C before harvesting by centrifugation at 4000 rpm for 20 min at 4 °C.

Small scale RNCs purification The final cell pellet was resuspended in lysis buffer (50 mM HEPES, 12 mM $Mg(OAc)_2$, 500 mM KOAc, 2 mM BME, 100 mg/ml lysozyme, pH 7.6) and lysed by 5 freeze/thaw cycles. The cell lysate was centrifuged at 13000 rpm for 25 mins at 4 °C in a bench-top centrifuge. A trace of DNAseI was added if the lysate was found to be glutinous. The supernatant was collected and loaded onto a 30 % (w/v) sucrose cushion (300 µl sample + 700 µl cushion) prepared in Buffer B. The cushion was then centrifuged at 120000 rpm for 30 mins at 4 °C in a TLA120.1 rotor (or 100,000 rpm for 90 mins in a TLA110 rotor). 200 µl from the top of the supernatant was carefully removed and precipitated by addition of 1 ml ice-cold acetone. The sample was centrifuged at 13000 rpm for 15 mins at

Methanol/chloroform precipitation All reagents were ice-cold. For 150 µl sample, 600 µl of methanol was added and mixed thoroughly before adding 150 µl of chloroform and mixing further by vortexing. 450 µl of ddH_2O was then added and mixed. The resulting mixture was put on ice for 2 mins before centrifugation for 5 mins in a bench-top centrifuge at top speed. After centrifugation, the mixture separated into two layers, an aqueous top layer and an organic bottom layer between which is a layer of precipitate containing the sample. The top layer was carefully removed without disturbing the disc. Another 650 µl of methanol was added and after mixing the sample was centrifuged for 5 mins in a bench-top centrifuge at top speed. All liquid was removed after centrifugation and pellet air-dried. The precipitated protein in the pellet was then resuspended in 20 µl of Tico buffer for further analysis.

5.1.15 Quantification of ribosome and RNC concentration by UV absorbance

The concentrations of ribosomes and the ribosomal content of RNC samples were determined by UV spectroscopy from the absorvance at 260 nm, A_{260}=1 in a 1 cm path-length corresponds to 24 nM of 70S ribosomes [257].The A_{260}/A_{280} ratio was used as a measure of the purity of the ribosome samples as this ratio reflects the relative proportion of RNA and proteins present [257], a pure ribosome sample should give a ratio between 1.9-2.0.

5.1.16 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), immunodetection (western blotting) and RNC occupancy determination

5.1.16.1 SDS-PAGE

Samples were mixed with LDS loading dye (Invitrogen) and boiled for 5 min before loading onto pre-cast 4-12 % (w/v) Bis-Tris gels (Invitrogen). The maximum sample volume are 15 and 25 µl for 15- and 10-well gels, respectively. The gels were ran at 200 V for 35 min in 1× MES running buffer (Invitrogen).

Gels of RNC and ribosome samples were typically silver-stained using the SilverQuest silver-staining kit (Invitrogen), with a staining procedure according to the manufacturer's instructions. Gels of isolated proteins were usually stained with Coomassie stain (0.025 % w/v coomassie brilliant blue, 40 % v/v ethanol, 10 % acetic acid) and destained using destain (40 % v/v ethanol, 10 % acetic acid).

5.1.16.2 Immunodetection - western blotting

Buffer compositions

- Transfer buffer: 6.05 g/l Tris, 14.4 g/l glycine, 20 % (v/v) methanol, 0.01 % (w/v) SDS.
- Tris-buffered Saline (TBS): 2.42 g/l Tris,11.69 g/l NaCl, pH 7.4.

- TBS-Tween (TBST): 0.05 % (v/v) Tween-20 in TBS

- Blocking buffer:

  Histidine-Tag detection: 1 % (w/v) casein in TBS

  Trigger-factor (TF) detection: 0.5 % (w/v) milk powder in TBST

  αSyn detection: 5 % (w/v) milk powder in TBST

  SecM detection: 5 % (w/v) bovine serum albumin (BSA) in TBST

- Antibodies:

  Histidine-Tag detection: Penta-His HRP conjugate (1:5000) (Qiagen) in TBST

  anti-TF primary antibody: rabbit anti-TF polyclonal antibody (1:2500) (Genscript) in blocking buffer

  anti-TF secondary antibody: Horseradish peroxidase conjugated anti-rabbit antibody (raised in goat) (1:1000) (New England Biolabs) in blocking buffer

  anti-αSyn primary antibody: mouse anti-αSyn (1:1000) in blocking buffer

  anti-αSyn secondary antibody: anti-mouse IgG (1:1000) in blocking buffer

  anti-SecM primary antibody: rabbit anti-SecM (1:10000) in 2 % (w/v) BSA in TBST

  anti-SecM secondary antibody: anti-rabbit (1:20000) in 2 % (w/v) BSA in TBST

- Chemiluminescence detection (ECL): SuperSignal West Pico chemiluminescence substrate (Pierce)

Western-blot procedure. The gel after SDS-PAGE was soaked in cold transfer buffer, together with six sponges, nitrocellulose membrane and 2 filter papers, before stacking into the transfer chamber in the following order: 3 sponges, 1 filter paper, gel, nitrocellulose membrane, 1 filter paper, 3 sponges (Figure 5.1), taking care to avoid air bubbles. The apparatus was assembled into the gel running tank. The inner chamber was filled with transfer buffer and the outer chamber was filled with water. The transfer was performed for 2 h at a current of 250 mA on ice. After transfer, the membrane was soaked in blocking buffer and incubated for 1 hour with shaking. The gel after transfer was stained with Coomassie stain to confirm the efficienty of the transfer. After blocking, the membrane

was incubated in primary antibody for 2 h or overnight, and then washed 4 times for 10 min in TBST. For anti-histidine detection, the membrane was incubated with ECL reagent for 5 min and the chemiluminescence detected using a Fujifilm LAS-1000 scanner with exposure times ranging from 30 s to 10 min depending on the intensity. For other detections, the membrane was first incubated in the secondary antibody for 1 hour at room temperature, followed by another washing stage using TBST (4×10 min), before ECL incubation and chemiluminescence detection as described above. The densities of the western blot signals were quantified with Image J software[258].

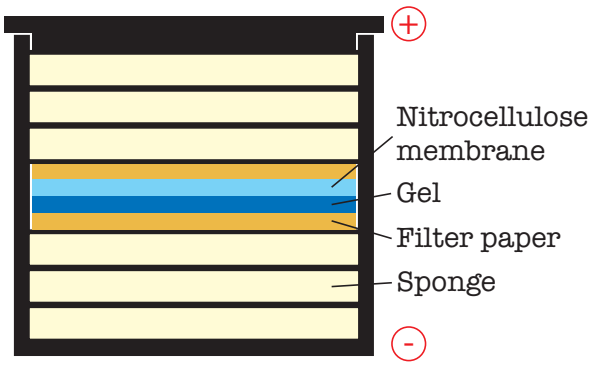

Figure 5.1: Schematic illustrating the stacking layers for western blot transfer.

5.1.16.3 Determination of nascent chain occupancy and trigger factor binding determination

To measure the NC occupancy, 7.5 pmol (for αSyn) or 10 pmol (for Dom5+110) of the RNC together with a series of the corresponding purified isolated protein standards 2 to 10 pmol) were run by SDS-PAGE, transferred onto the western blot membrane and probed for the his-tag. Densitometric analysis of the resulting bands enabled a standard curve based upon the isolated protein standards to be generated, which was then used to determine the NC occupancy. The quantification of TF binding was performed in a similar manner, except that the standards used were purified TF, and these were loaded in amounts ranging from 1 to 10 pmol.

5.1.17 Analytical gel-filtration chromatography

The conformational properties of the isolated L7/L12 protein were evaluated using analytical gel filtration performed on an Akta FPLC (GE healthcare) using a Superdex 75 10/300 GL column. The column was pre-equilibrated with phosphate buffer and the elution was monitored by following the absorbance at 230 nm.

5.1.18 Fluorescence titration measurements

Fluorescence measurements were performed in a 1 cm path length quartz cell at room temperature. Measurements were made using a FluoroMax-4 spectrofluorometer (HORIBA Jobin Yvon) with excitation and emission slit widths set at 1 nm and an integration time of 0.1 s. A scan speed of 100 nm/s was used and the emission spectra were averaged after 4 independent acquisitions.

The binding to the lanthanide binding tag was evaluated by titrating Dom5-LBT and Dom5 (each at a concentration of 1 μM, in 10 mM HEPES, 100 NaCl, pH 7.5) with a range of 0-25 molar equivalents of terbium (Tb^{3+}) (prepared in water). The fluorescence of the Tb^{3+} was excited at a wavelength of 280 nm and emission spectra were collected over 300-600 nm (0.5 nm increments). The fluorescence maxima at 489 nm and 544 nm were used for analysis.

5.2 NMR sample preparation and conditions

5.2.1 NMR sample conditions

Sample, labelling, concentration, buffer	Experiment
Fraction 1, ^{15}N, 160 µM, Tico, H_2O	isotropic N-H splittings, IPAP-^{15}N-HSQC
Fraction 2, ^{15}N, 240 µM, Tico, H_2O	isotropic N-H splittings, IPAP-^{15}N-HSQC
Fraction 1,^{15}N, 110 µM, Tico, H_2O +phage, ~10 mg ml^{-1}, 5.4 Hz ^{2}H splittings	phage-aligned N-H splittings, IPAP-^{15}N-HSQC
Fraction 2,^{15}N, 170 µM, Tico, H_2O +phage, ~10 mg ml^{-1}, 3.7 Hz ^{2}H splittings	phage-aligned N-H splittings, IPAP-^{15}N-HSQC
Fraction 2, ^{13}C/^{15}N, 500 µM, phosphate, H_2O	Triple-resonance assignment experiments
Fraction 2, ^{13}C/^{15}N, 250 µM, Tico, H_2O	^{15}N relaxation experiments
Fraction 2, ^{13}C/^{15}N, 20 µM, phosphate, H_2O	C^α-H^α sensitivity tests IPAP-C^α-HSQC IPAP-BEST-HN(CO)CA IPAP-(HA)CA(CO)NH
Fraction 2, ^{13}C/^{15}N, 20 µM, Tico, D_2O	isotropic C^α-H^α splittings IPAP-C^α-HSQC
Fraction 2,^{13}C/^{15}N, 20 µM, Tico, D_2O +phage, ~7.5 mg ml^{-1}, 2.4 Hz ^{2}H splittings	phage-aligned C^α-H^α splittings, IPAP-C^α-HSQC
Fraction 2,^{13}C/^{15}N, 250 µM, Tico, H_2O +Otting, 5 %, 30 Hz ^{2}H splittings	Otting-aligned N-H splittings, ^{15}N-HSQC/TROSY
Fraction 2, ^{13}C/^{15}N, 20 µM, Tico, H_2O	isotropic C'-C^α splittings/sensitivity test, IPAP-HNCO
70S ribosome, ^{15}N, 10 µM, Tico, H_2O	isotropic N-H splittings, HSQC/TROSY
70S ribosome, ^{15}N, 10 µM, Tico, H_2O +phage, 12.3 Hz ^{2}H splittings	phage-aligned N-H splittings, HSQC/TROSY
70S ribosome, ^{13}C/^{15}N, 10 µM, Tico, D_2O	isotropic C^α-H^α splittings, IPAP-C^α-HSQC
70S ribosome, ^{13}C/^{15}N, 10 µM, Tico, H_2O +Otting, 5 %, 18 Hz ^{2}H splittings, 3 %, 15 Hz	Otting-aligned N-H splittings (attempted)
70S ribosome, ^{13}C/^{15}N, 10 µM, Tico, H_2O	isotropic N-C'/H-C' splittings, IPAP-E.COSY-^{15}N-HSQC-TROSY
70S ribosome, ^{15}N, 10 µM, Tico, H_2O +phage, ~17.5 mg ml^{-1}, 12.9 Hz ^{2}H splittings	phage-aligned C'-C^α splittings, IPAP-HNCO (attempted)

Table 5.9: NMR sample conditions of isolated and ribosome-bound L7/L12. Tico buffer: 10 mM HEPES, 30 mM NH_4Cl, 12 mM $MgCl_2$, 5 mM EDTA, 2 mM BME, 0.05 % NaN_3, protease inhibitors, pH 7.6. Phosphate buffer: 50 mM Na_2HPO_4, 100 mM KCl, 5 mM EDTA, 0.05 % NaN_3, pH 7.6. All experiments recorded at 25 °C.

Sample, labelling, concentration, buffer	Experiment
Dom5,^{15}N, 100 µM, Tico, H_2O	isotropic N-H splittings, IPAP-^{15}N-HSQC
Dom5,^{15}N, 100 µM, Tico, H_2O +phage, 12.8 Hz ^{2}H splittings	phage-aligned N-H splittings, IPAP-^{15}N-HSQC
Dom5-LBT,^{15}N, 100 µM, Tico, H_2O	isotropic N-H splittings, IPAP-^{15}N-HSQC
Dom5-LBT,^{15}N, 100 µM, Tico, H_2O	LBT-aligned N-H splittings, IPAP-^{15}N-HSQC
Dom5+110 RNC, ^{15}N, 9.8 µM, Tico, H_2O	isotropic N-H splittings, ^{15}N HSQC/TROSY
V82M-αSyn, ^{15}N, 100 µM, Tico, H_2O	isotropic N-H splittings, ^{15}N HSQC/TROSY
V82M-αSyn,^{15}N, 100 µM, Tico, H_2O +phage, ∼10 mg ml^{-1}, 6.4 Hz ^{2}H splittings	phage-aligned N-H splittings, ^{15}N HSQC/TROSY
V82M-αSyn-RNC, ^{15}N, 5.25 µM, Tico, H_2O	isotropic splittings, ^{15}N HSQC/TROSY
V82M-αSyn-RNC, ^{15}N, 4.4 µM, Tico, H_2O +phage, ∼18 mg ml^{-1}, 13.8 Hz ^{2}H splittings	phage-aligned N-H splittings, ^{15}N HSQC/TROSY
WT-αSyn-RNC, ^{15}N, 4.3 µM, Tico, H_2O	^{15}N-HSQC
V82M-αSyn, ^{15}N, 5 µM, Tico, H_2O	testing interaction with ribosome, HSQC/SOFAST-HMQC
V82M-αSyn, ^{15}N, 5 µM, 140mM NH_4Cl, H_2O	testing interaction with ribosome, HSQC/SOFAST-HMQC
V82M-αSyn, ^{15}N, 5 µM, 100 mM Glu+Asp, H_2O	testing interaction with ribosome, HSQC/SOFAST-HMQC
V82M-αSyn, ^{15}N, 5 µM, Tico, H_2O +70S ribosome, unlabelled, 5 µM	testing interaction with ribosome, HSQC/SOFAST-HMQC
V82M-αSyn, ^{15}N, 5 µM, 140mM salt, H_2O +70S ribosome, unlabelled, 5 µM	testing interaction with ribosome, HSQC/SOFAST-HMQC
V82M-αSyn, ^{15}N, 5 µM, Glu+Asp, H_2O +70S ribosome, unlabelled, 5 µM	testing interaction with ribosome, HSQC/SOFAST-HMQC

Table 5.10: NMR sample conditions for RNCs and isolated proteins. All spectra were recorded at 25 °C except for αSyn spectra, which were recorded at 4 °.

5.2.2 Alignment media preparation for RDCs experiments

5.2.2.1 Preparation of Pf1 phage

The phage alignment medium used in this work was prepared using bacteriophage Pf1 (ASLA Biotech Ltd.). The phage were centrifuged at 95000 rpm for 1 h at 5 °C in a Beckman TLA-120.1 rotor and resuspended in Tico buffer. The procedure was repeated twice to buffer exchange the phage into Tico buffer. The concentration of phage was determined from the UV absorbance at 270 nm, with $\epsilon = 2.25\,\mathrm{mg\ ml^{-1}cm^{-1}}$. The deuterium splitting was measured from the 1D ^{2}H NMR spectrum to confirm the concentration of phage in the NMR sample and to verify the homogeneity of the aligned liquid-crystal phase.

5.2.2.2 Preparation of Otting medium

The Otting medium used in this work was prepared from $C_{12}E_5$ (pentaethylene glycol monododecyl ether) and *n*-hexanol, with r = 0.85 (where r is the molar ratio of $C_{12}E_5$ to hexanol) according to [134]. A stock solution of the medium at 10 % (w/w) $C_{12}E_5$ was prepared by first mixing $C_{12}E_5$ and buffer, and then adding n-hexanol stepwise with vortexing between each addition. This stock solution was then diluted in a 1:1 ratio with the protein/ribosome sample to give a final $C_{12}E_5$ concentration of 5 % (w/w).

5.3 NMR data collection

Except for the ^{15}N relaxation measurements, all NMR experiments were recorded on a Bruker Avance III 700 MHz spectrometer equipped with a cryogenically-cooled inverse triple-resonance probehead (CP-TCI) and running Topspin 2.1. ^{15}N relaxation experiments were recorded on a Bruker Avance III 500 MHz equipped with a room-temperature inverse triple-resonance probehead (TXI) and running Topspin 2.1. All non-standard NMR sequences were coded by Dr John Kirkpatrick, UCL.

5.3.1 Ribosome monitoring

5.3.1.1 1D proton spectra

1D proton spectra were acquired using either pre-saturation or excitation sculpting [259] for water suppression, with typical acquisition times of 100-200 ms. Pre-saturation 1D spectra were acquired with a recycle delay of 2 s, and pre-saturation field strengths of 50-100 Hz. Excitation-sculpting spectra were acquired with a recycle delay of 2 s, and using 2 ms square water-selective 180 ° pulses.

5.3.1.2 SOFAST-HMQC spectra

SOFAST-^{15}N-HMQC spectra [260] were acquired with typical acquisition times of 50 ms and 30 ms in ^{1}H and ^{15}N, respectively, and recycle delays of 50-100 ms. GARP ^{15}N decoupling [261] during acquisition was applied with a field strength of 0.7 kHz. 120 ° PC-9 [262] and 180 ° ReBURP [263] pulses centred at 8.2 ppm and with approximate bandwidths of 3.1 ppm were used for amide proton magnetisation excitation and refocusing, respectively. ^{13}C decoupling during t_1 (where necessary) was applied with a 500-µs smoothed-CHIRP adiabatic pulse [264] centred at 115 ppm with an 80-kHz frequency sweep.

5.3.1.3 Labelling experiments

The ^{15}N-edited labelling experiment was run using a modified ^{15}N-SOFAST-HMQC sequence with pre-saturation added for extra water suppression. The ^{15}N-filtered experiment was run with the same sequence, but with the phase-cycle of the receiver inverted to reject ^{15}N-labelled magnetisation.

5.3.1.4 Diffusion experiemnts

All diffusion experiments were run as stimulated gradient echoes [265], where the diffusion delay and the gradient lengths were fixed, and the gradient strengths were varied. ^{15}N- and ^{13}C-X-stimulated echoes were recorded according to [266]. Bipolar gradients (smoothed-square shape) were used for encoding and decoding the spatially dependent phase. Water

suppression was achieved using water saturation (50-100 Hz field strength) during both the recycle and diffusion delays. Typical parameters were: encode/decode gradient length, δ = 4 ms, diffusion delay, Δ = 100 ms, gradient strengths: 5 %, 50 %, 95 % (expressed as percentages of the maximum gradient strength). The maximum gradient strength was 55.57 G/cm.

5.3.2 N-H splittings

Backbone amide N-H splittings were measured using either the IPAP or HSQC/TROSY approaches. ^{15}N HSQC spectra [267] were recorded using phase-cycling for coherence order selection, and WATERGATE with flipback [268] for water suppression. ^{15}N HSQC-TROSY spectra [269] were recorded with phase-cycling for coherence order selection and 3-9-19 WATERGATE [270] for water suppression. The IPAP spectra were acquired in an interleaved fashion with an IPAP-modified ^{15}N-HSQC sequence [157]. The in-phase and anti-phase preparation periods were of equal length ($1/2J_{\mathrm{NH}}$ = 5.4 ms).

5.3.3 C^{α}-H^{α} splittings

Measurement of C^{α}-H^{α} splitting was tested using three sequences. The IPAP-BEST-HN(CO)CA sequence was modified from the BEST-HN(CO)CA sequence (see assignment section below) by addition of a C^{α}-H^{α} IPAP element of length $1/2J_{C^{\alpha}H^{\alpha}}$ = 3.6 ms prior to the C^{α} indirect evolution period. The C^{α} acquisition time was set to 9 ms. The IPAP-(HA)CA(CO)NH sequence was modified from the (HA)CA(CO)NH sequence (see assignment section) by addition of the same IPAP element prior to the C^{α} indirect evolution period. The constant-time C^{α} evolution period was set to 6.4 ms. C^{α}-H^{α} IPAP elements were implemented using C^{α}-selective ReBURP refocusing pulses (22 ppm bandwidth centred at 59 ppm).

The IPAP-C^{α}-HSQC sequence was adapted in a similar manner from a C^{α}-selective ^{13}C HSQC sequence. The C^{α}-selective ^{13}C HSQC sequence was developed from a standard gradient-selected, sensitivity-enhanced ^{13}C HSQC sequence. C^{α}-selectivity was achieved

using ReBURP/IBurp1 pulses (22 ppm bandwidth centred at 59 ppm) in the first INEPT and at the end of t_1 [263]. Selective C'/C^{β}-decoupling was implemented using a doubly-cosine-modulated IBurp1 pulse during t1. The cosine-modulation and pulse length were adjusted to give the following three inversion bands:

- C': 32 ppm bandwidth centred at 175.5 ppm, Thr C^{β}: 32 ppm bandwidth centred at 85 ppm, Upfield C^{β}: 32 ppm bandwidth centred at 30.5 ppm
- $^{13}C^{\alpha}$-HSQC and IPAP-C^{α}-HSQC spectra were acquired with ^{13}C acquisition times of 25-30 ms.

5.3.4 N-C' and H-C' splittings

N-C' and H-C' splittings were measured using an IPAP-E.COSY-^{15}N-HSQC-TROSY sequence, adapted from the ^{15}N-HSQC-TROSY sequence by addition of an N-C' IPAP element prior to the t_1 indirect evolution period. The in-phase and anti-phase preparation elements were of equal length ($1/2J_{NC'}$ = 33 ms). C'-selective inversion during the IPAP element was achieved using a Q3 pulse [271] centred at 176 ppm with 22 ppm bandwidth. C^{α}-decoupling during t1 was implemented with a Q3 pulse centred at 56 ppm with 80 ppm bandwidth. The proton and ^{15}N acquisition times were set to 76 ms and 62 ms, respectively.

5.3.5 C^{α}-C' splittings

C^{α}-C' splittings were measured using an IPAP-BEST-HNCO sequence modified from the BEST-HNCO sequence (see below) by addition of a C^{α}-C' IPAP element of length $1/2J_{C^{\alpha}C'}$ = 9 ms prior to the C' indirect evolution time. The C' acquisition time was set to 22-24 ms.

5.3.6 Assignment experiments

Data for backbone assignment of L7/L12 were acquired using the BEST implementations (band-selective excitation short-transient) of the standard HNCO, HNCA, HN(CO)CA, HNCACB, HN(CO)CACB and HN(CA)CO triple-resonance assignment experiments [224].

BEST experiments were acquired with proton acquisition times of 50 ms, and recycle delays of 250-350 ms. ^{15}N decoupling during acquisition was applied with a field strength of 0.83 kHz. All H^N-selective pulses were centred at 8.2 ppm with approximate bandwidths of 3-3.5 ppm. H^N-selective refocusing pulses were applied as ReBURP pulses. H^N-selective 90 ° pulses were applied as PC-9 pulses in the first INEPT and as EBurp2 pulses in the back-transfer sensitivity-enhancement block. Proton decoupling during the ^{13}C and ^{15}N indirect evolution periods was achieved with broadband-inversion pulses [272].

The H^α resonances were assigned using HA(CACO)NH and HA(CA)NH sequences modified from the corresponding standard HBHA(CBCACO)NH and HBHA(CBCA)NH sequences [273]. H^α/C^α assignments were transferred to the constant-time ^{13}C HSQC spectrum [274].

5.3.7 Relaxation experiments

Backbone ^{15}N relaxation rates ($R_1$and $R_{1\rho}$) were measured on isolated L7/L12 using established methods [275, 276]. The water signal was preserved using selective water pulses (2 ms sinc shape) and weak bipolar gradients during the indirect chemical shift evolution time to maintain the magnetisation along z. In the R_1 sequence, N-H cross-relaxation pathways were suppressed by application of amide-selective IBurp1 pulses (centred at 8.2 ppm) at intervals of 10 ms during the relaxation delay [277, 278]. In the $R_{1\rho}$ sequence, cross-relaxation was suppressed by application of random-phase proton CW during the nitrogen spin-lock [279], and magnetization was aligned with the 1.5 kHz spin-lock field [280]. Backbone $\{^1H\}^{15}N$ steady-state heteronuclear NOEs were measured using the standard method [275, 276]. Water magnetisation was preserved in the reference spectrum as described above. Saturation of the amide proton magnetisation was achieved using a 5 s train of high-power 120 ° pulses applied at 5 ms intervals. The reference and saturated spectra were recorded in an interleaved fashion. To ensure full recovery of the water magnetisation at the start of each increment of the reference experiment, a long recycle delay of 15 s was used [281].

5.4 NMR data processing and analysis

NMR data were processed using the NMRPipe [282] and Azara software packages [283] Spectral analysis, peak picking, assignment and curve-fitting for extraction of relaxation rates were performed using the CcpNmr Analysis program [284]. Signal-to-noise ratios were calculated as the ratio of the peak height to the RMS noise. The RMS noise was calculated (Azara) as the standard deviation of the data points within a signal-free region. The linewidths (full-width at half-height) were measured in CcpNmr Analysis. The uncertainties in peak positions were calculated according to Equation 2.4 (see section 2.2.3.4 in the main text). Uncertainties in R_1 and $R_{1\rho}$ relaxation rate constants were calculated using the boot-strapping Monte-Carlo function in CcpNmr Analysis. Uncertainties in the $\{^1H\}^{15}N$-heteronuclear NOE values were calculated according to the noise levels in the reference and saturated sub-spectra.

5.4.1 Diffusion experiments with less than four gradient strengths

Diffusion coefficients for experiments recorded with 4 or fewer gradient strengths were calculated as described in section 2.2.1.5 in the main text.

The uncertainty were calculated according to:

$$\delta D = k\sqrt{\left(\frac{\delta I_1}{I_1}\right)^2 + \left(\frac{\delta I_2}{I_2}\right)^2}$$

where I is the integrated signal intensity and δI is calculated from a set of integrals over signal-free regions of the same width as the integrated signal region using the following equation:

$$\delta I = \sqrt{\frac{1}{N - 1.5} \cdot \sum_i (x_i - \overline{x})^2}$$

where x_i are the individual integrals, $\overline{x}$ is the mean integral and N is the number of integrals. k is given by:

$$k = \frac{1}{\left(\Delta - \frac{\delta}{3} - \frac{\tau}{2}\right)(\gamma\delta)^2\left(G_2^2 - G_1^2\right)}$$

5.4.2 Diffusion experiments with more than four gradient strengths

The diffusion coefficients for experiments recorded with more than 4 gradient strengths were calculated using the build-in function 'T1guide, in Topspin, in which the signal region is selected interactively and the resulting integrated intensities fitted to Equation 1.3 to extract the diffusion coefficient.

5.4.3 Relaxation data analysis

The measured $R_{1\rho}$ relaxation rates were converted to R_2 relaxation rates according to:

$$R_{1\rho} = R_1 \cdot \cos^2\theta + R_2 \cdot \sin^2\theta$$

θ is the tip-angle of the effective spin-lock field, given by:

$$\theta = \tan^{-1}(\omega_1/\Omega)$$

where ω_1 is the spin-lock field strength and Ω is the ^{15}N offset, respectively.

The ^{15}N heteronuclear relaxation rates were analysed according to the model-free formalism [210] using the program TENSOR2 [285]. The analysis proceeds in two stages. First, the global diffusion tensor is derived from the R_2/R_1 ratios for residues showing negligible internal motion. For this purpose, all residues with a heteronuclear NOE of less than 0.6 were excluded. In addition, residues showing a residual exchange contribution to their transverse relaxation were excluded. These residues were identified according the exclusion criterion:

$$\frac{\langle T_2\rangle - T_2}{\langle T_2\rangle} - \frac{T_1 - \langle T_1\rangle}{\langle T_1\rangle} > 1.5 \times \mathrm{SD} \tag{5.1}$$

where T_1 and T_2 are the inverses of R_1 and R_2, respectively, and SD is the standard deviation of the quantity on the left-hand side of the inequality.

In the second stage, the anisotropic diffusion tensor is used to fit the relaxation rates for

all residues to the model-free spectral density function, yielding order parameters and, where necessary, internal correlation times.

Estimates of the uncertainties in the extracted parameters were derived using the built-in Monte-Carlo function together with the estimated uncertainties in the measured relaxation rates.

5.4.4 RDC data analysis

L7/L12 alignment tensors and back-calculated RDCs were calculated by fitting the measured RDCs to the L7/L12 NMR structure (PDB 1RQU) using SVD as implemented in the program PALES [213, 286]. As discussed in the main text, the extracted RDC values were filtered according to spectral overlap, internal flexibility and exchange contributions before PALES fitting.

Uncertainties in the alignment tensor parameters were derived using the built-in PALES Monte-Carlo module mcDc. In this method, noise is added to experimental RDCs. The amount of noise added is adjusted iteratively such that 80 % of the 1000 back-calculated data-sets contain no back-calculated RDCs that deviate from the experimental values by more than twice the standard deviation of the added noise [287].

5.4.5 Structure refinement

Structure refinement of the C-terminal domain of the NMR structure of L7/L12 (PDB 1RQU) was performed using Xplor-NIH [215] using a protocol based on that described in [214]. A single iteration was performed with the NCS weight set to 100000 for residues in secondary-structure elements. The RDC data were included as SANI restraints, using a smooth-square potential with well-width set to the experimental uncertainty and a force constant of 1. Only RDCs from residues in secondary-structure elements were included as restraints. D_{a} and R values for the SANI potential were taken as the values derived from PALES fitting to the NMR structure. The orientation of the alignment tensor was allowed to float. Pairwise RMSDs between the refined and starting structures were calculated in PyMOL [288] over

the backbone C^{α}, N and C' atoms.

Appendix A

Additional plots

A.1 1D ^{1}H spectra of isolated L7/L12

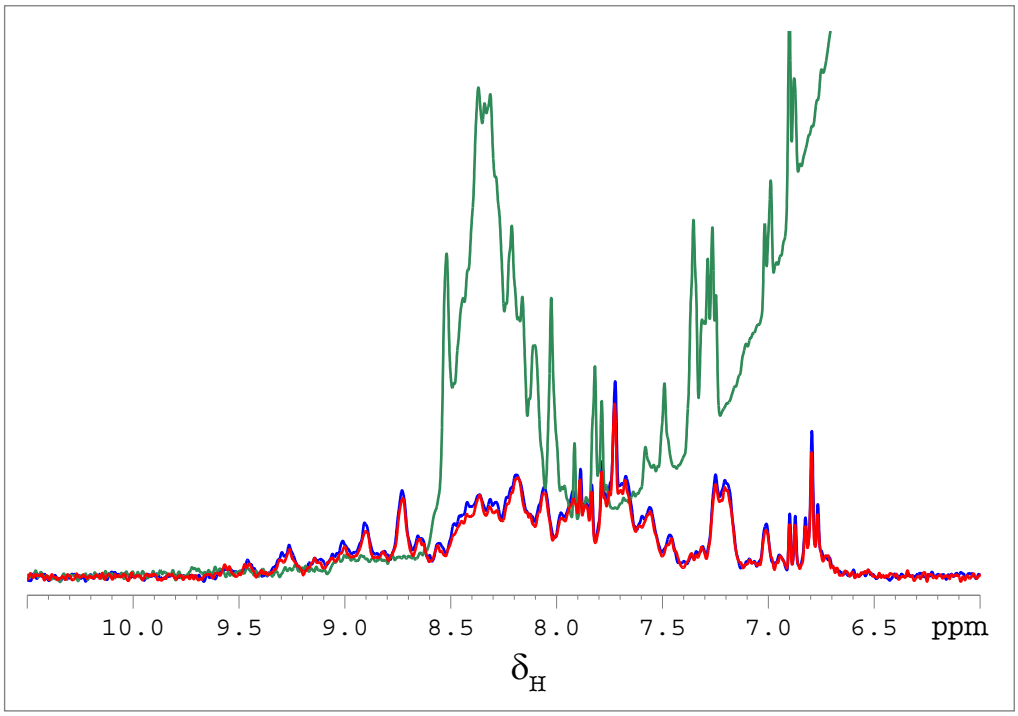

Figure A.1: 1D ^{1}H spectra of isolated L7/L12 samples from the analytical gel-filtration experiments. Blue: before denaturation, green: denatured in 8M urea, red: after refolding. The curved base-line of the green spectrum is the result of the strong urea signal.

A.2 Diffusion plots for 70S ribosome samples

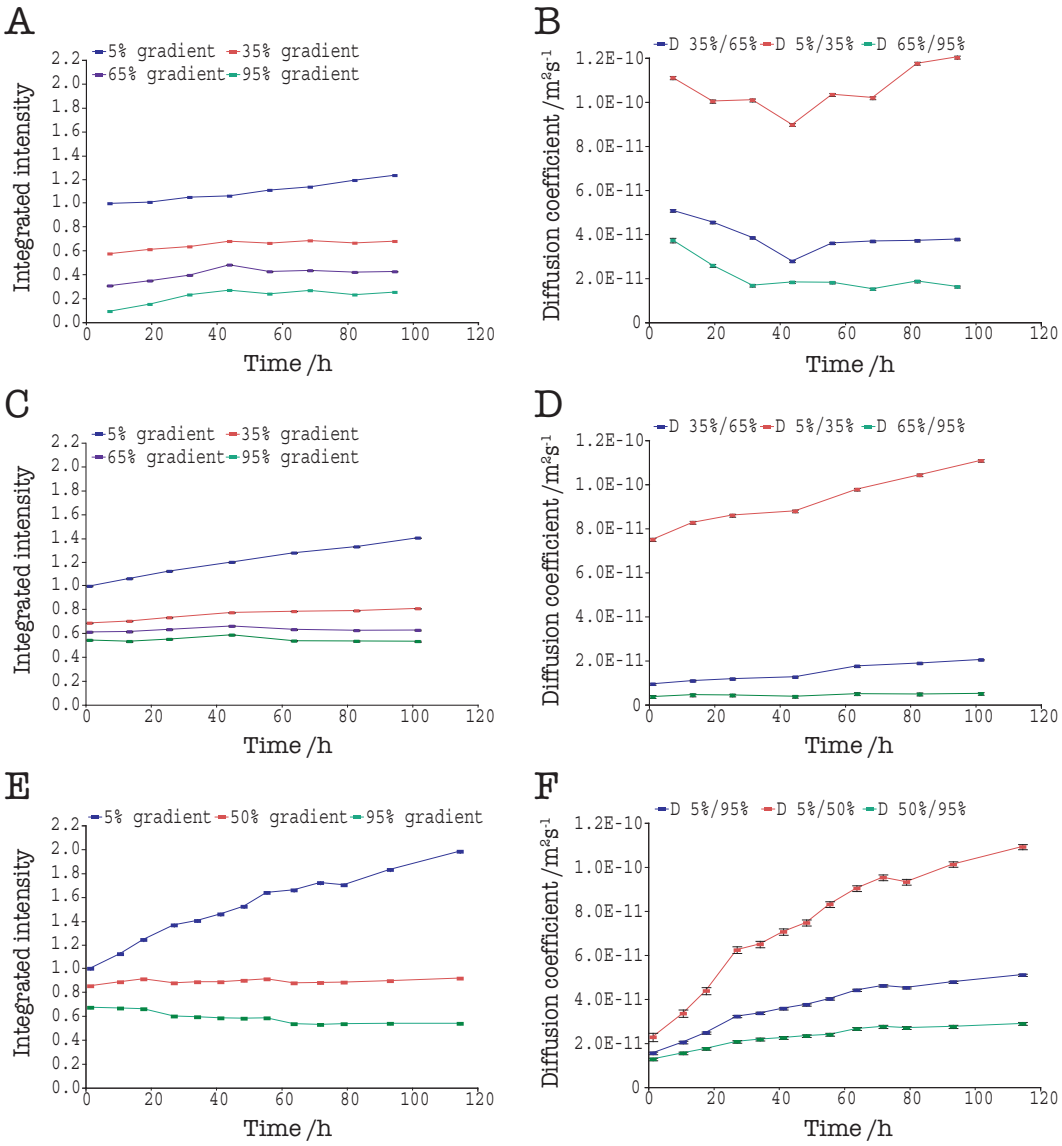

Figure A.2: Plot of the integrated signal intensity (A, C and E) and the calculated diffusion coefficient (B, D and F) for isotropic sample of ^{15}N-labelled ribosomes (A and B), the aligned sample of ^{15}N-labelled ribosomes (C and D) and the isotropic sample of ^{15}N,^{13}C-labelled ribosomes (E and F).

A.3 N-H RDCs of isolated and ribosome-bound L7/L12

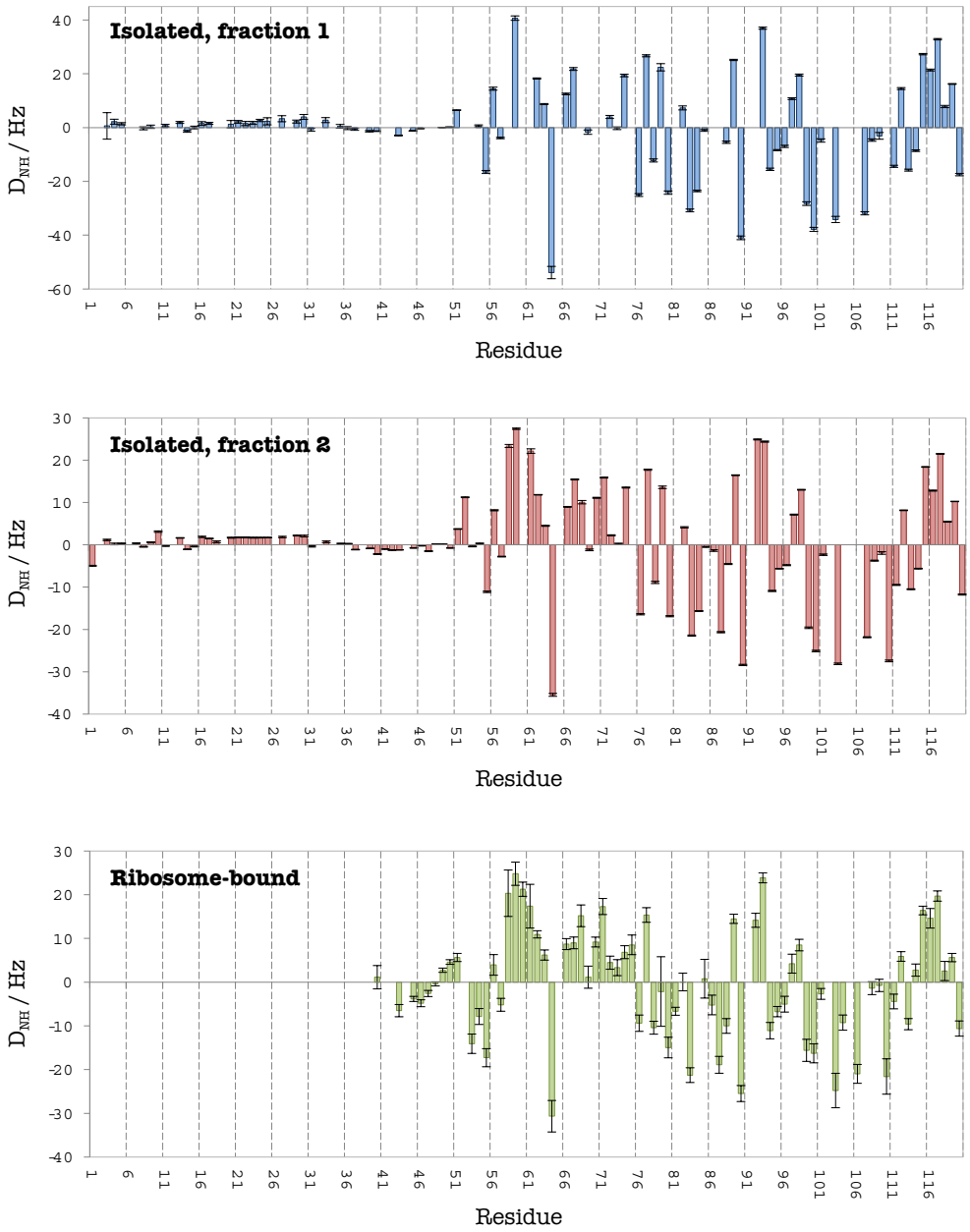

Figure A.3: N-H RDCs measured for fraction 1 and fraction 2 of isolated L7/L12 and for ribosome-bound L7/L12, with phage alignment.

A.4 ^{15}N relaxation rates (R_1, R_2 and heteronuclear NOE) of isolated L7/L12

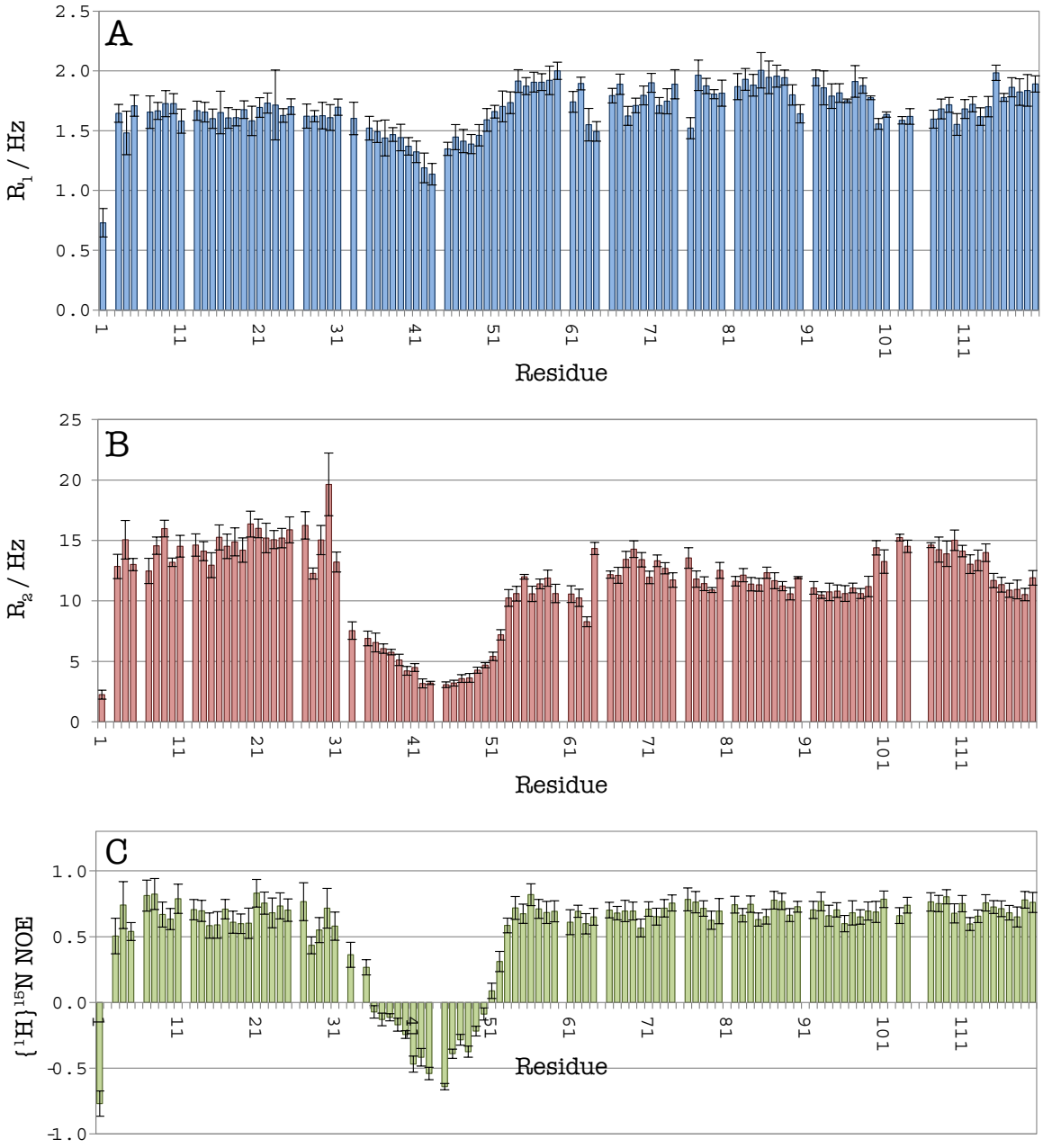

Figure A.4: ^{15}N relaxation rates (R_1, R_2 and heteronuclear NOE) of isolated L7/L12.

A.5 C^{α}-H^{α} RDCs measured in phage and N-H RDCs measured in Otting medium for isolated L7/L12

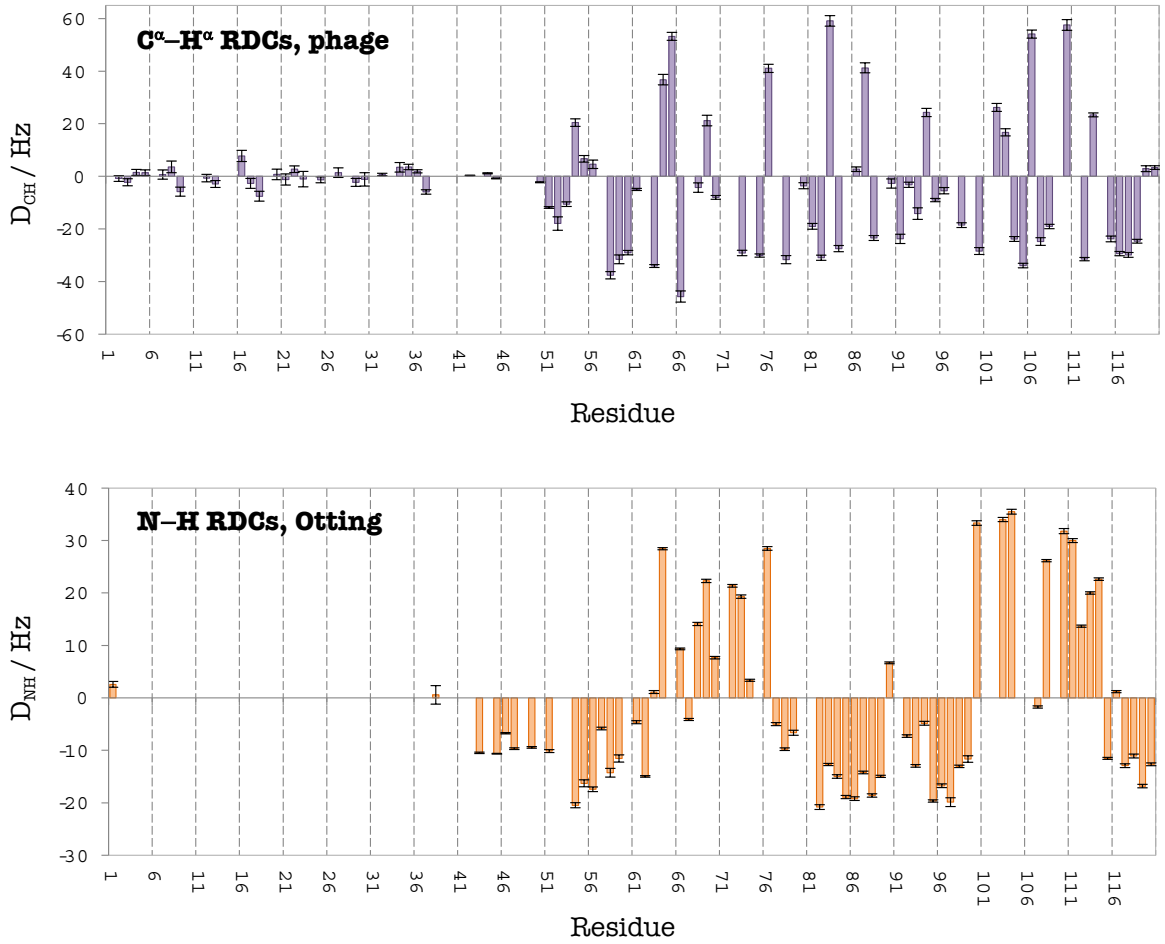

Figure A.5: C^{α}-H^{α} RDCs measured in phage and N-H RDCs measured in Otting medium for isolated L7/L12.

A.6 N-H assignments of the isolated L7/L12 protein

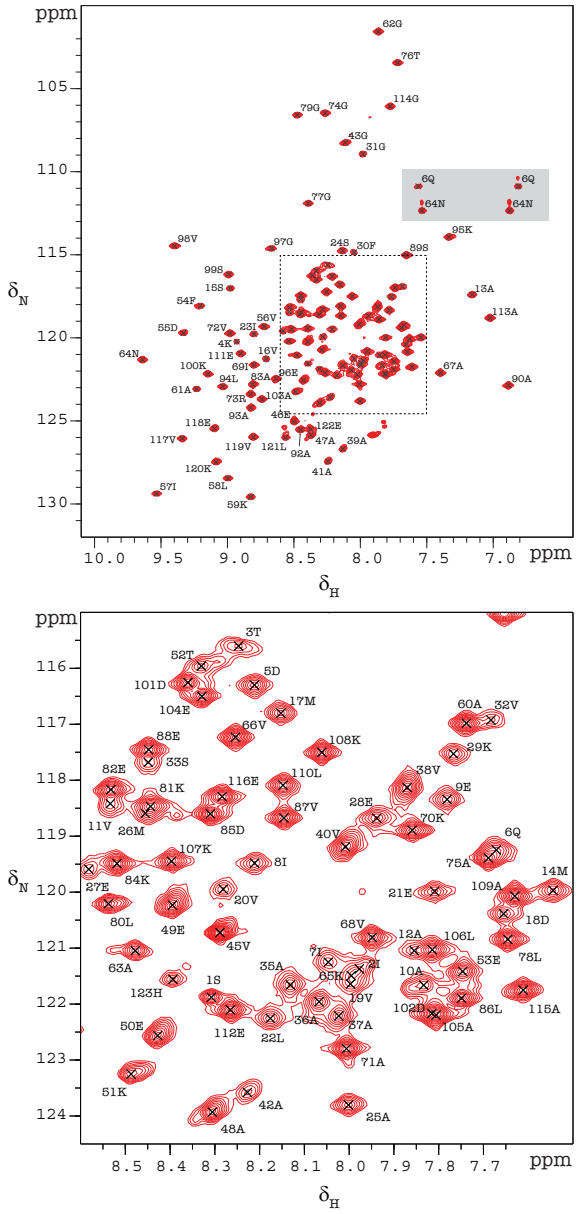

Figure A.6: ^{15}N HSQC spectrum of isolated L7/L12 showing peak assignments. The bottom panel shows an expanded view of the crowded central region for clarity. Peaks in the grey box correspond to the sidechain NH_2 resonances from the asparagine and glutamine residues.

A.7 IPAP-E.COSY-^{15}N-HSQC-TROSY spectra of ^{13}C,^{15}N-labelled 70S ribosomes

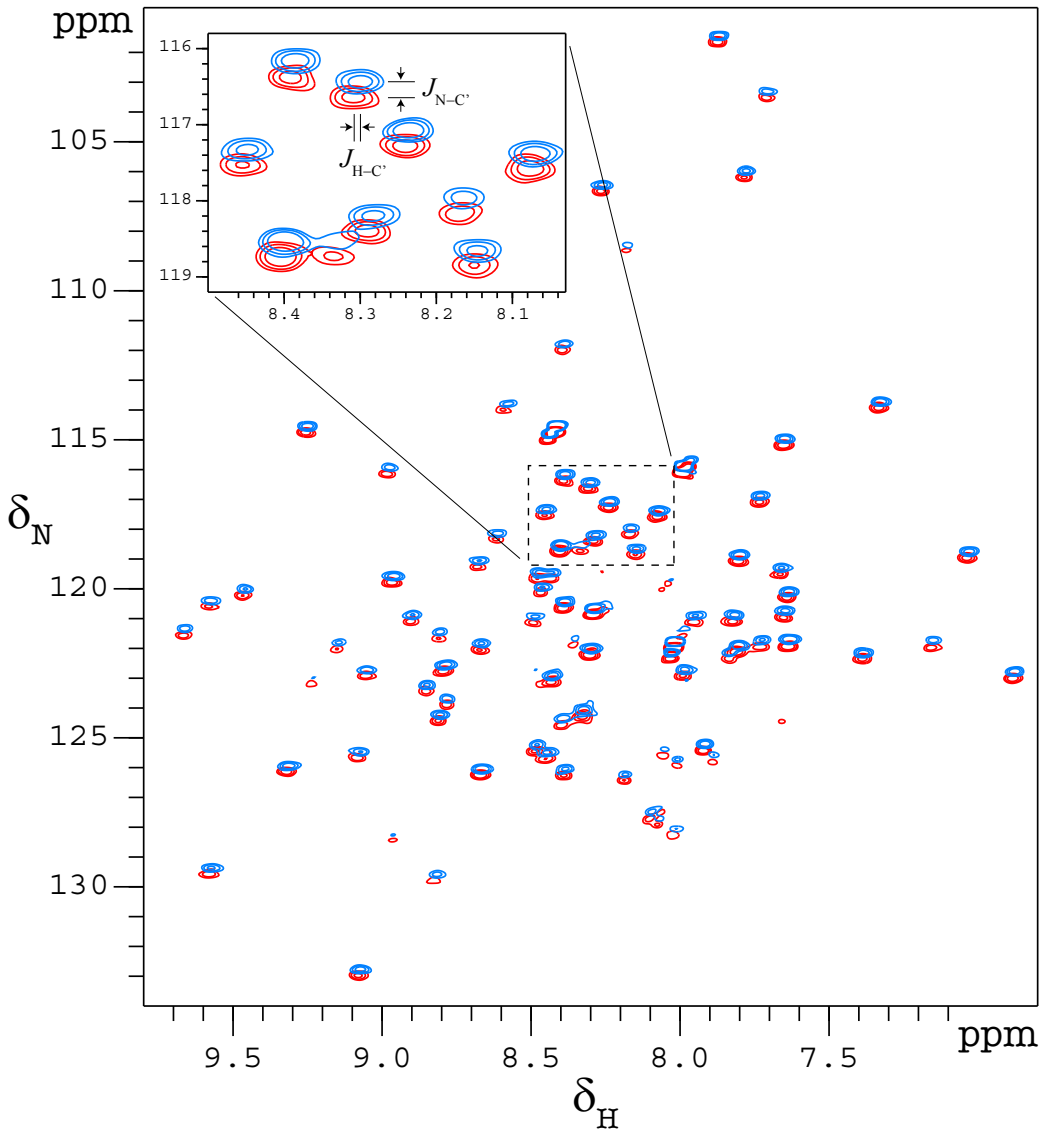

Figure A.7: IPAP-E.COSY-^{15}N-HSQC-TROSY spectra of ^{13}C,^{15}N-labelled 70S ribosomes.

A.8 N-C' and H-C' J couplings of ribosome-bound L7/L12 and C'-C$^{\alpha}$ J couplings of isolated L7/L12

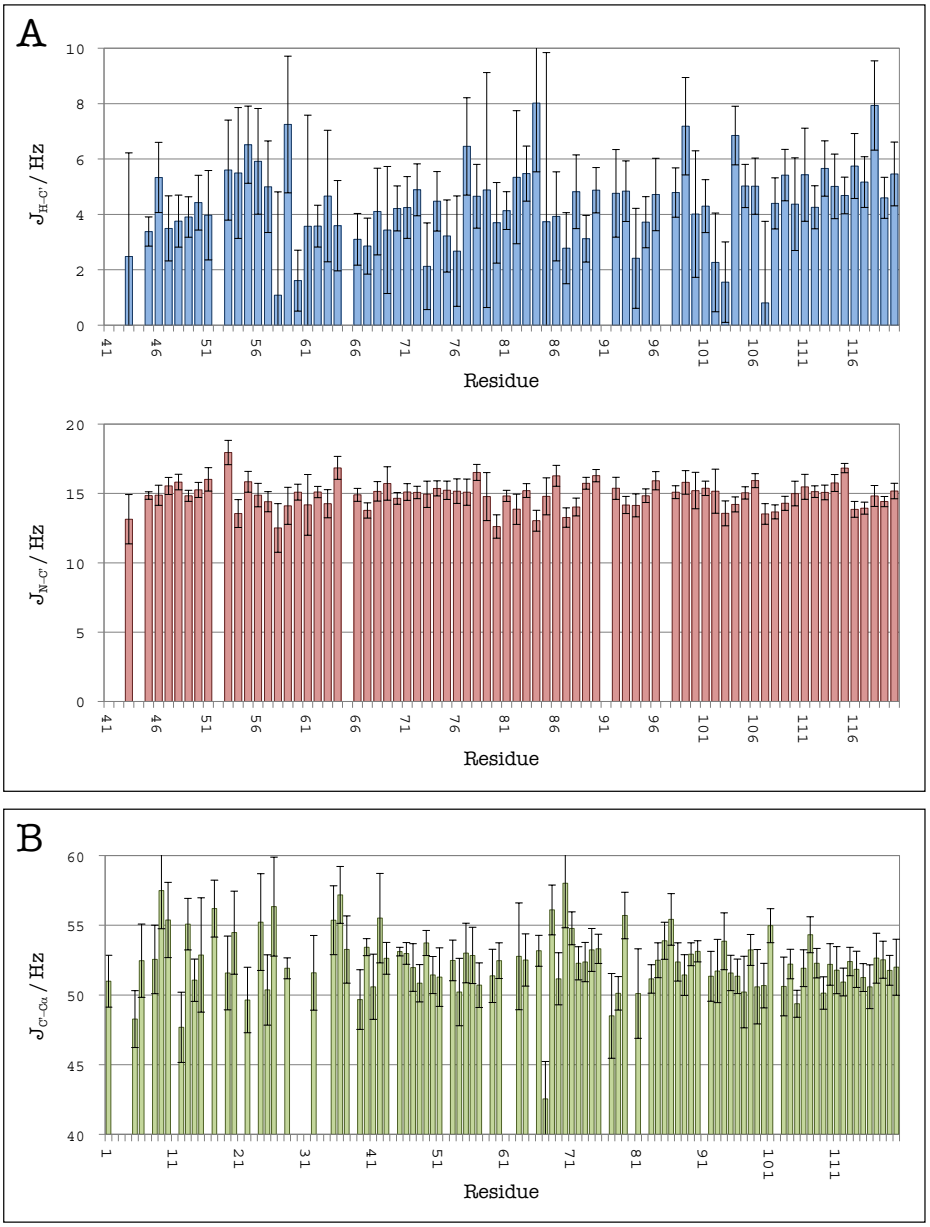

Figure A.8: **A:** N-C' and H-C' J couplings of ribosome-bound L7/L12, measured on 10 μM 70S ribosomes. **B:** C'-C$^{\alpha}$ J couplings of isolated L7/L12, measured on 20 μM isolated L7/L12.

Appendix B

Protein and DNA sequences

B.1 L7/L12 sequences

DNA sequence

ATGTCTATCA CTAAAGATCA AATCATTGAA GCAGTTGCAG CTATGTCTGT AATGGACGTT
GTAGAACTGA TCTCTGCAAT GGAAGAAAAA TTCGGTGTTT CCGCTGCTGC TGCTGTAGCT
GTAGCTGCTG GCCCGGTTGA AGCTGCTGAA GAAAAAACTG AATTCGACGT AATTCTGAAA
GCTGCTGGCG CTAACAAAGT TGCTGTTATC AAAGCAGTAC GTGGCGCAAC TGGCCTGGGT
CTGAAAGAAG CTAAAGACCT GGTAGAATCT GCACCGGCTG CTCTGAAAGA AGGCGTGAGC
AAAGACGACG CAGAAGCACT GAAAAAAGCT CTGGAAGAAG CTGGCGCTGA AGTTGAAGTT
AAACTCGAGC ACCACCACCA CCACCACTGA

Amino acid sequence

MSITKDQIIE AVAAMSVMDV VELISAMEEK FGVSAAAAVA VAAGPVEAAE EKTEFDVILK
AAGANKVAVI KAVRGATGLG LKEAKDLVES APAALKEGVS KDDAEALKKA LEEAGAEVEV
KLEHHHHHH

Within the amino acid sequence, L7/L12 residues are highlighted in red.

B.2 ddFLN$_{646\text{-}838}$ RNC and related protein sequences

DNA sequence

ATGCATCACC ATCACCATCA CGCTAGCAAA CCAGCCCCAT CCGCTGAACA CTCTTATGCT
GAAGGTGAAG GTTTAGTCAA AGTATTTGAT AATGCCCCAG CTGAATTCAC TATTTTCGCC
GTTGACACTA AAGGTGTTGC TCGTACCGAT GGTGGTGATC CATTTGAAGT CGCTATCAAT
GGTCCAGATG GTTTAGTCGT TGATGCCAAA GTTACCGATA ACAATGACGG TACTTATGGT
GTTGTCTATG ATGCCCCAGT TGAAGGTAAC TACAATGTTA ATGTCACCCT CCGTGGTAAT
CCAATCAAAA ATATGCCAAT CGATGTCAAA TGCATTGAAG GTGCCAATGG TGAAGATTCA
TCATTCGGTT CATTCACTTT TACCGTCGCT GCTAAAAATA AGAAAGGTGA AGTTAAAACC
TATGGTGGTG ATAAATTTGA AGTCTCTATC ACTGGTCCAG CTGAAGAAAT CACTCTCGAT
GCTATTGATA ACCAAGATGG TACTTATACT GCCGCTTACT CTTTAGTTGG TAATGGTCGT
TTCTCAACTG GTGTCAAATT AAACGGTAAA CACATTGAAG GTTCTCCATT CAAACAAGTA
CTTGGTAACA CTAGTGAATT CTTCAGCACG CCCGTCTGGA TAAGCCAGGC GCAAGGCATC
CGTGCTGGCC CTTAA

Amino acid sequence

MHHHHHHASK PAPSAEHSYA EGEGLVKVFD NAPAEFTIFA VDTKGVARTD GGDPFEVAIN
GPDGLVVDAK VTDNNDGTYG VVYDAPVEGN YNVNVTLRGN PIKNMPIDVK CIEGANGEDS
SFGSFTFTVA AKNKKGEVKT YGGDKFEVSI TGPAEEITLD AIDNQDGTYT AAYSLVGNGR
FSTGVKLNGK HIEGSPFKQV LGNTSEFFST PVWISQAQGI RAGP

Within the amino acid sequence, domain 5 residues are highlighted in red, domain 6 residues are highlighted in green and the SecM residues are highlighted in blue. The isolated Dom5 and Dom5+6 protein sequence correspond to the red and red+green sequence respectively.

B.3 α-Synuclein sequences

DNA sequence

ATGCATCACC ATCACCATCA CGCTAGCATG GATGTATTCA TGAAAGGACT TTCAAAGGCC
AAGGAGGGAG TTGTGGCTGC TGCTGAGAAA ACCAAACAGG GTGTGGCAGA AGCAGCAGGA
AAGACAAAAG AGGGTGTTCT CTATGTAGGC TCCAAAACCA AGGAGGGAGT GGTGCATGGT
GTGGCAACAG TGGCTGAGAA GACCAAAGAG CAAGTGACAA ATGTTGGAGG AGCAGTGGTG
ACGGGTGTGA CAGCAGTAGC CCAGAAGACA GTGGAGGGAG CAGGGAGCAT TGCAGCAGCC
ACTGGCTTTG TCAAAAAGGA CCAGTTGGGC AAGAATGAAG AAGGAGCCCC ACAGGAAGGA
ATTCTGGAAG ATATGCCTGT GGATCCTGAC AATGAGGCTT ATGAAATGCC TTCTGAGGAA
GGGTATCAAG ACTACGAACC TGAAGCCGGT ACCACTAGTG AATTCTTCAG CACGCCCGTC
TGGATAAGCC AGGCGCAAGG CATCCGTGCT GGCCCTTAAC CATGGACCTA ACAACAATAA

Amino acid sequence

MHHHHHHASM DVFMKGLSKA KEGVVAAAEK TKQGVAEAAG KTKEGVLYVG SKTKEGVVHG
VATVAEKTKE QVTNVGGAVV TGVTAVAQKT VEGAGSIAAA TGFVKKDQLG KNEEGAPQEG
ILEDMPVDPD NEAYEMPSEE GYQDYEPEAG TTSEFFSTPV WISQAQGIRA GP

Within the amino acid sequence, α-Synuclein residues are highlighted in red and the SecM residues are highlighted in blue. The residues highlighted in pink arise from cloning.

Bibliography

[1] Arnold, R.J. & Reilly, J.P. Observation of*Escherichia coli*Ribosomal Proteins and Their Posttranslational Modifications by Mass Spectrometry. *Analytical Biochemistry* **269**, 105–112 (1999).

[2] Dunkle, J.A. & Cate, J.H.D. Ribosome structure and dynamics during translocation and termination. *Annual Review of Biophysics* **39**, 227–244 (2010).

[3] Ban, N., Nissen, P., Hansen, J., Moore, P.B. & Steitz, T.A. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905–920 (2000).

[4] Naganuma, T., Nomura, N., Yao, M., Mochizuki, M., Uchiumi, T. & Tanaka, I. Structural Basis for Translation Factor Recruitment to the Eukaryotic/Archaeal Ribosomes. *The Journal of Biological Chemistry* **285**, 4747–4756 (2010).

[5] Dunkle, J.A., Xiong, L., Mankin, A.S. & Cate, J.H.D. Structures of the *Escherichia coli* ribosome with antibiotics bound near the peptidyl transferase center explain spectra of drug action. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 17152–17157 (2010).

[6] Ben-Shem, A., Jenner, L., Yusupova, G. & Yusupov, M. Crystal structure of the eukaryotic ribosome. *Science* **330**, 1203–1209 (2010).

[7] Rabl, J., Leibundgut, M., Ataide, S.F., Haag, A. & Ban, N. Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science* **331**, 730–736 (2011).

[8] Gao, Y.G., Selmer, M., Dunham, C.M., Weixlbaumer, A., Kelley, A.C. & Ramakrishnan, V. The structure of the ribosome with elongation factor G trapped in the posttranslocational state. *Science* **326**, 694–699 (2009).

[9] Schmeing, T.M., Huang, K.S., Strobel, S.A. & Steitz, T.A. An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. *Nature* **438**, 520–524 (2005).

[10] Gao, H., Sengupta, J., Valle, M., Korostelev, A., Eswar, N., Stagg, S.M., Van Roey, P., Agrawal, R.K., Harvey, S.C., Sali, A., Chapman, M.S. & Frank, J. Study of the structural dynamics of the E coli 70S ribosome using real-space refinement. *Cell* **113**, 789–801 (2003).

[11] Julián, P., Milon, P., Agirrezabala, X., Lasso, G., Gil, D., Rodnina, M.V. & Valle, M. The Cryo-EM Structure of a Complete 30S Translation Initiation Complex from *Escherichia coli*. *PLoS biology* **9**, e1001095 (2011).

[12] Villa, E., Sengupta, J., Trabuco, L.G., LeBarron, J., Baxter, W.T., Shaikh, T.R., Grassucci, R.A., Nissen, P., Ehrenberg, M., Schulten, K. & Frank, J. Ribosome-induced changes in elongation factor Tu conformation control GTP hydrolysis. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 1063–1068 (2009).

[13] Schuette, J.C., Murphy, F.V., Kelley, A.C., Weir, J.R., Giesebrecht, J., Connell, S.R., Loerke, J., Mielke, T., Zhang, W., Penczek, P.A., Ramakrishnan, V. & Spahn, C.M.T. GTPase activation of elongation factor EF-Tu by the ribosome during decoding. *The EMBO journal* **28**, 755–765 (2009).

[14] Selmer, M., Dunham, C.M., Murphy, F.V., Weixlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R. & Ramakrishnan, V. Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935–1942 (2006).

[15] Armache, J.P., Jarasch, A., Anger, A.M., Villa, E., Becker, T., Bhushan, S., Jossinet, F., Habeck, M., Dindar, G., Franckenberg, S., Marquez, V., Mielke, T., Thomm, M., Berninghausen, O., Beatrix, B., Söding, J., Westhof, E., Wilson, D.N. & Beckmann, R. Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5 Å resolution. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 19748–19753 (2010).

[16] Becker, T., Armache, J.P., Jarasch, A., Anger, A.M., Villa, E., Sieber, H., Motaal, B.A., Mielke, T., Berninghausen, O. & Beckmann, R. Structure of the no-go mRNA decay complex Dom34-Hbs1 bound to a stalled 80S ribosome. *Nature Structural & Molecular Biology* **18**, 715–720 (2011).

[17] Nissen, P., Hansen, J., Ban, N., Moore, P.B. & Steitz, T.A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930 (2000).

[18] Nirenberg, M. & Leder, P. RNA codewords and protein synthesis. *Science* **145**, 1399–1407 (1964).

[19] Monro, R.E. Catalysis of peptide bond formation by 50S ribosomal subunits from *Escherichia coli*. *J Mol Biol* **26**, 147–151 (1967).

[20] Kosolapov, A. & Deutsch, C. Tertiary interactions within the ribosomal exit tunnel. *Nature Structural & Molecular Biology* **16**, 405–411 (2009).

[21] Schuwirth, B.S., Borovinskaya, M.A., Hau, C.W., Zhang, W., Vila-Sanjurjo, A., Holton, J.M. & Cate, J.H.D. Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **310**, 827–834 (2005).

[22] Bocharov, E.V. From Structure and Dynamics of Protein L7/L12 to Molecular Switching in Ribosome. *The Journal of Biological Chemistry* **279**, 17697–17706 (2004).

[23] Jenner, L.B., Demeshkina, N., Yusupova, G. & Yusupov, M. Structural aspects of messenger RNA reading frame maintenance by the ribosome. *Nature Structural & Molecular Biology* **17**, 555–560 (2010).

[24] Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H. & Noller, H.F. Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292**, 883–896 (2001).

[25] Seidelt, B., Innis, C.A., Wilson, D.N., Gartmann, M., Armache, J.P., Villa, E., Trabuco, L.G., Becker, T., Mielke, T., Schulten, K., Steitz, T.A. & Beckmann, R. Structural insight into nascent polypeptide chain-mediated translational stalling. *Science* **326**, 1412–1415 (2009).

[26] Schmeing, T.M. & Ramakrishnan, V. What recent ribosome structures have revealed about the mechanism of translation. *Nature* **461**, 1234–1242 (2009).

[27] Wahl, M.C. & Möller, W. Structure and Function of the Acidic Ribosomal Stalk Proteins. *Current Protein & Peptide Science* **3**, 93–106 (2002).

[28] Mohr, D., Wintermeyer, W. & Rodnina, M.V. GTPase activation of elongation factors Tu and G on the ribosome. *Biochemistry* **41**, 12520–12528 (2002).

[29] Diaconu, M., Kothe, U., Schlunzen, F., Fischer, N., Harms, J., Tonevitsky, A., Stark, H., Rodnina, M. & Wahl, M. Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. *Cell* **121**, 991–1004 (2005).

[30] Moore, P.B. How Should We Think About the Ribosome? *Annual Review of Biophysics* **41**, 1–19 (2012).

[31] Mitra, K. & Frank, J. Ribosome dynamics: insights from atomic structure modeling into cryo-electron microscopy maps. *Annual Review of Biophysics and Biomolecular Structure* **35**, 299–317 (2006).

[32] Petrov, A., Kornberg, G., O'Leary, S., Tsai, A., Uemura, S. & Puglisi, J.D. Dynamics of the translational machinery. *Current opinion in chemical biology* **21**, 137–145 (2011).

[33] Frank, J. & Gonzalez, R.L. Structure and dynamics of a processive Brownian motor: the translating ribosome. *Annu Rev Biochem* **79**, 381–412 (2010).

[34] Christodoulou, J., Larsson, G., Fucini, P., Connell, S.R., Pertinhez, T.A., Hanson, C.L., Redfield, C., Nierhaus, K.H., Robinson, C.V., Schleucher, J. & Dobson, C.M. Heteronuclear NMR investigations of dynamic regions of intact *Escherichia coli* ribosomes. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 10949–10954 (2004).

[35] Voorhees, R.M., Schmeing, T.M., Kelley, A.C. & Ramakrishnan, V. The mechanism for activation of GTP hydrolysis on the ribosome. *Science* **330**, 835–838 (2010).

[36] Lee, T.H., Blanchard, S.C., Kim, H.D., Puglisi, J.D. & Chu, S. The role of fluctuations in tRNA selection by the ribosome. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 13661–13665 (2007).

[37] Agirrezabala, X., Lei, J., Brunelle, J.L., Ortiz-Meoz, R.F., Green, R. & Frank, J. Visualization of the hybrid state of tRNA binding promoted by spontaneous ratcheting of the ribosome. *Molecular Cell* **32**, 190–197 (2008).

[38] Connell, S.R., Takemoto, C., Wilson, D.N., Wang, H., Murayama, K., Terada, T., Shirouzu, M., Rost, M., Schüler, M., Giesebrecht, J., Dabrowski, M., Mielke, T., Fucini, P., Yokoyama, S. & Spahn, C.M.T. Structural basis for interaction of the ribosome with the switch regions of GTP-bound elongation factors. *Molecular Cell* **25**, 751–764 (2007).

[39] Julián, P., Konevega, A.L., Scheres, S.H.W., Lázaro, M., Gil, D., Wintermeyer, W., Rodnina, M.V. & Valle, M. Structure of ratcheted ribosomes with tRNAs in hybrid states. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 16924–16927 (2008).

[40] Valle, M., Zavialov, A., Sengupta, J., Rawat, U., Ehrenberg, M. & Frank, J. Locking and unlocking of ribosomal motions. *Cell* **114**, 123–134 (2003).

[41] Frank, J., Gao, H., Sengupta, J., Gao, N. & Taylor, D.J. The process of mRNA-tRNA translocation. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 19671–19678 (2007).

[42] Agrawal, R.K., Heagle, A.B., Penczek, P., Grassucci, R.A. & Frank, J. EF-G-dependent GTP hydrolysis induces translocation accompanied by large conformational changes in the 70S ribosome. *Nature structural biology* **6**, 643–647 (1999).

[43] Zhang, W., Dunkle, J.A. & Cate, J.H.D. Structures of the Ribosome in Intermediate States of Ratcheting. *Science* **325**, 1014–1017 (2009).

[44] Berk, V., Zhang, W., Pai, R.D., Cate, J.H.D. & Cate, J.H.D. Structural basis for mRNA and tRNA positioning on the ribosome. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 15830–15834 (2006).

[45] Gabashvili, I.S., Agrawal, R.K., Spahn, C.M., Grassucci, R.A., Svergun, D.I., Frank, J. & Penczek, P. Solution structure of the *E. coli* 70S ribosome at 11.5 Å resolution. *Cell* **100**, 537–549 (2000).

[46] Korostelev, A., Asahara, H., Lancaster, L., Laurberg, M., Hirschi, A., Zhu, J., Trakhanov, S., Scott, W.G. & Noller, H.F. Crystal structure of a translation termination complex formed with release factor RF2. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 19684–19689 (2008).

[47] Laurberg, M., Asahara, H., Korostelev, A., Zhu, J., Trakhanov, S. & Noller, H.F. Structural basis for translation termination on the 70S ribosome. *Nature* **454**, 852–857 (2008).

[48] Weixlbaumer, A., Jin, H., Neubauer, C., Voorhees, R.M., Petry, S., Kelley, A.C. & Ramakrishnan, V. Insights into translational termination from the structure of RF2 bound to the ribosome. *Science* **322**, 953–956 (2008).

[49] Fei, J., Bronson, J.E., Hofman, J.M., Srinivas, R.L., Wiggins, C.H. & Gonzalez, R.L. Allosteric collaboration between elongation factor G and the ribosomal L1 stalk directs tRNA movements during translation. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 15702–15707 (2009).

[50] Becker, T., Bhushan, S., Jarasch, A., Armache, J.P., Funes, S., Jossinet, F., Gumbart, J., Mielke, T., Berninghausen, O., Schulten, K., Westhof, E., Gilmore, R., Mandon, E.C. & Beckmann, R. Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome. *Science* **326**, 1369–1373 (2009).

[51] Bhushan, S., Hoffmann, T., Seidelt, B., Frauenfeld, J., Mielke, T., Berninghausen, O., Wilson, D.N. & Beckmann, R. SecM-stalled ribosomes adopt an altered geometry at the peptidyl transferase center. *PLoS biology* **9**, e1000581 (2011).

[52] Bhushan, S., Gartmann, M., Halic, M., Armache, J.P., Jarasch, A., Mielke, T., Berninghausen, O., Wilson, D.N. & Beckmann, R. alpha-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. *Nature Structural & Molecular Biology* **17**, 313–317 (2010).

[53] Munro, J.B., Sanbonmatsu, K.Y., Spahn, C.M.T. & Blanchard, S.C. Navigating the ribosome's metastable energy landscape. *Trends in Biochemical Sciences* **34**, 390–400 (2009).

[54] Whitford, P.C., Geggier, P., Altman, R.B., Blanchard, S.C., Onuchic, J.N. & Sanbonmatsu, K.Y. Accommodation of aminoacyl-tRNA into the ribosome involves reversible excursions along multiple pathways. *RNA* **16**, 1196–1204 (2010).

[55] Fulle, S. & Gohlke, H. Statics of the ribosomal exit tunnel: implications for cotranslational peptide folding, elongation regulation, and antibiotics binding. *J Mol Biol* **387**, 502–517 (2009).

[56] Petrone, P.M., Snow, C.D., Lucent, D. & Pande, V.S. Side-chain recognition and gating in the ribosome exit tunnel. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 16549–16554 (2008).

[57] Elcock, A.H. Molecular simulations of cotranslational protein folding: fragment stabilities, folding cooperativity, and trapping in the ribosome. *PLoS Comput Biol* **2**, e98 (2006).

[58] O'Brien, E.P., Hsu, S.T.D., Christodoulou, J., Vendruscolo, M. & Dobson, C.M. Transient Tertiary Structure Formation within the Ribosome Exit Port. *Journal of the American Chemical Society* **132**, 16928–16937 (2010).

[59] O'Brien, E.P., Christodoulou, J., Vendruscolo, M. & Dobson, C.M. New Scenarios of Protein Folding Can Occur on the Ribosome. *Journal of the American Chemical Society* **133**, 513–526 (2011).

[60] Dobson, C.M. Protein folding and misfolding. *Nature* **426**, 884–890 (2003).

[61] Anfinsen, C.B. Principles that govern the folding of protein chains. *Science* **181**, 223–230 (1973).

[62] Gershenson, A. & Gierasch, L.M. Protein folding in the cell: challenges and progress. *Current opinion in chemical biology* **21**, 32–41 (2011).

[63] Xu, Z., Horwich, A.L. & Sigler, P.B. The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. *Nature* **388**, 741–750 (1997).

[64] Elliott, P.R., Pei, X.Y., Dafforn, T.R. & Lomas, D.A. Topography of a 2.0 Å structure of 1-antitrypsin reveals targets for rational drug design to prevent conformational disease. *Protein Science* **9**, 1274–1281 (2000).

[65] Eliezer, D., Kutluay, E., Bussell Jr, R. & Browne, G. Conformational properties of α-synuclein in its free and lipid-associated states. *J Mol Biol* **307**, 1061–1073 (2001).

[66] Hartl, F.U., Bracher, A. & Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* **475**, 324–332 (2011).

[67] Chiti, F. & Dobson, C.M. Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* **75**, 333–366 (2006).

[68] Levinthal, C. Are there pathways for protein folding? *Journal of Medical Physics* **65**, 44–45 (1968).

[69] Fersht, A. Characterizing transition states in protein folding: an essential step in the puzzle. *Current opinion in chemical biology* **5**, 79–84 (1995).

[70] Villali, J. & Kern, D. Choreographing an enzyme's dance. *Curr opin Chem Biol* **14**, 636–643 (2010).

[71] Englander, S. & Mayne, L. Protein folding studied using hydrogen-exchange labeling and two-dimensional NMR. *Annu Rev of Biophys Biomol Struct* **21**, 243–265 (1992).

[72] Baldwin, R. Pulsed H/D-exchange studies of folding intermediates. *Current opinion in chemical biology* **3**, 84–91 (1993).

[73] Schwaiger, I., Schleicher, M., Noegel, A.A. & Rief, M. The folding pathway of a fast-folding immunoglobulin domain revealed by single-molecule mechanical experiments. *EMBO Rep* **6**, 46–51 (2005).

[74] Hartl, F.U. & Hayer-Hartl, M. Converging concepts of protein folding in vitro and in vivo. *Nature Structural & Molecular Biology* **16**, 574–581 (2009).

[75] Kubelka, J., Chiu, T.K., Davies, D.R., Eaton, W.A. & Hofrichter, J. Sub-microsecond protein folding. *J Mol Biol* **359**, 546–553 (2006).

[76] Kaiser, C.M., Goldman, D.H., Chodera, J.D., Tinoco, I. & Bustamante, C. The ribosome modulates nascent protein folding. *Science* **334**, 1723–1727 (2011).

[77] Plaxco, K.W., Simons, K.T. & Baker, D. Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* **277**, 985–994 (1998).

[78] Geierhaas, C.D., Paci, E., Vendruscolo, M. & Clarke, J. Comparison of the transition states for folding of two Ig-like proteins from different superfamilies. *J Mol Biol* **343**, 1111–1123 (2004).

[79] Gardino, A.K., Villali, J., Kivenson, A., Lei, M., Liu, C.F., Steindel, P., Eisenmesser, E.Z., Labeikovsky, W., Wolf-Watz, M., Clarkson, M.W. & Kern, D. Transient non-native hydrogen bonds promote activation of a signaling protein. *Cell* **139**, 1109–1118 (2009).

[80] Brockwell, D.J. & Radford, S.E. Intermediates: ubiquitous species on folding energy landscapes? *Current opinion in chemical biology* **17**, 30–37 (2007).

[81] Zhang, G., Hubalewska, M. & Ignatova, Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nature Structural & Molecular Biology* **16**, 274–280 (2009).

[82] Zhang, G. & Ignatova, Z. Generic algorithm to predict the speed of translational elongation: implications for protein biogenesis. *PLoS ONE* **4**, e5036 (2009).

[83] Ellis, R.J. Macromolecular crowding: obvious but underappreciated. *Trends in Biochemical Sciences* **26**, 597–604 (2001).

[84] Tsao, D. & Dokholyan, N.V. Macromolecular crowding induces polypeptide compaction and decreases folding cooperativity. *Physical Chemistry Chemical Physics* **12**, 3491–3500 (2010).

[85] Elcock, A.H. Models of macromolecular crowding effects and the need for quantitative comparisons with experiment. *Current opinion in chemical biology* **20**, 196–206 (2010).

[86] Hong, J. & Gierasch, L.M. Macromolecular crowding remodels the energy landscape of a protein by favoring a more compact unfolded state. *Journal of the American Chemical Society* **132**, 10445–10452 (2010).

[87] Wang, Y., He, H. & Li, S. Effect of Ficoll 70 on thermal stability and structure of creatine kinase. *Biochemistry (Moscow)* **75**, 648–654 (2010).

[88] van den Berg, B., Ellis, R.J. & Dobson, C.M. Effects of macromolecular crowding on protein folding and aggregation. *The EMBO journal* **18**, 6927–6933 (1999).

[89] Mittal, J. & Best, R.B. Dependence of protein folding stability and dynamics on the density and composition of macromolecular crowders. *Biophysical Journal* **98**, 315–320 (2010).

[90] Brandt, F., Etchells, S.A., Ortiz, J.O., Elcock, A.H., Hartl, F.U. & Baumeister, W. The native 3D organization of bacterial polysomes. *Cell* **136**, 261–271 (2009).

[91] Hartl, F.U. & Hayer-Hartl, M. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295**, 1852–1858 (2002).

[92] Frydman, J. Folding of newly translated proteins in vivo: the role of molecular chaperones. *Annu Rev Biochem* **70**, 603–647 (2001).

[93] Hoffmann, A., Bukau, B. & Kramer, G. Structure and function of the molecular chaperone Trigger Factor. *Biochimica et biophysica acta* **1803**, 650–661 (2010).

[94] Ferbitz, L., Maier, T., Patzelt, H., Bukau, B., Deuerling, E. & Ban, N. Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* **431**, 590–596 (2004).

[95] Merz, F., Boehringer, D., Schaffitzel, C., Preissler, S., Hoffmann, A., Maier, T., Rutkowska, A., Lozza, J., Ban, N., Bukau, B. & Deuerling, E. Molecular mechanism and structure of Trigger Factor bound to the translating ribosome. *The EMBO journal* **27**, 1622–1632 (2008).

[96] Lakshmipathy, S.K., Tomic, S., Kaiser, C.M., Chang, H.C., Genevaux, P., Georgopoulos, C., Barral, J.M., Johnson, A.E., Hartl, F.U. & Etchells, S.A. Identification of nascent chain interaction sites on trigger factor. *The Journal of Biological Chemistry* **282**, 12186–12193 (2007).

[97] Cabrita, L.D., Dobson, C.M. & Christodoulou, J. Protein folding on the ribosome. *Current opinion in chemical biology* **20**, 33–45 (2010).

[98] Fedorov, A.N. Cotranslational Protein Folding. *The Journal of Biological Chemistry* **272**, 32715–32718 (1997).

[99] Komar, A., Kommer, A., Krasheninnikov, I.A. & Spirin, A.S. Cotranslational folding of globin. *The Journal of Biological Chemistry* **272**, 10646–10651 (1997).

[100] Hardesty, B. & Kramer, G. *Folding of a nascent peptide on the ribosome*, volume 66. Prog Nucleic Acid Res Mol Biol (2000).

[101] Kolb, V.A. Cotranslational protein folding. *Molecular Biology* **35**, 584–590 (2001).

[102] Johnson, A.E. The co-translational folding and interactions of nascent protein chains: a new approach using fluorescence resonance energy transfer. *FEBS Letters* **579**, 916–920 (2005).

[103] Land, A., Zonneveld, D. & Braakman, I. Folding of HIV-1 envelope glycoprotein involves extensive isomerization of disulfide bonds and conformation-dependent leader peptide cleavage. *FASEB J* **17**, 1058–1067 (2003).

[104] Chen, W., Helenius, J., Braakman, I. & Helenius, A. Cotranslational folding and calnexin binding during glycoprotein synthesis. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 6229–6233 (1995).

[105] Makeyev, E.V., Kolb, V.A. & Spirin, A.S. Enzymatic activity of the ribosome-bound nascent polypeptide. *FEBS Letters* **378**, 166–170 (1996).

[106] Kudlicki, W., Chirgwin, J., Kramer, G. & Hardesty, B. Folding of an enzyme into an active conformation while bound as peptidyl-tRNA to the ribosome. *Biochemistry* **34**, 14284–14287 (1995).

[107] Nicola, A.V., Chen, W. & Helenius, A. Co-translational folding of an alphavirus capsid protein in the cytosol of living cells. *Nat Cell Biol* **1**, 341–345 (1999).

[108] Tsalkova, T., Odom, O.W., Kramer, G. & Hardesty, B. Different conformations of nascent peptides on ribosomes. *J Mol Biol* **278**, 713–723 (1998).

[109] Clark, P.L. & King, J. A newly synthesized, ribosome-bound polypeptide chain adopts conformations dissimilar from early in vitro refolding intermediates. *The Journal of Biological Chemistry* **276**, 25411–25420 (2001).

[110] Kleizen, B., van Vlijmen, T., de Jonge, H.R. & Braakman, I. Folding of CFTR is predominantly cotranslational. *Molecular Cell* **20**, 277–287 (2005).

[111] Frydman, J., Erdjument-Bromage, H., Tempst, P. & Hartl, F.U. Co-translational domain folding as the structural basis for the rapid de novo folding of firefly luciferase. *Nature structural biology* **6**, 697–705 (1999).

[112] Woolhead, C.A., McCormick, P.J. & Johnson, A.E. Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. *Cell* **116**, 725–736 (2004).

[113] Ellis, J.P., Culviner, P.H. & Cavagnero, S. Confined dynamics of a ribosome-bound nascent globin: Cone angle analysis of fluorescence depolarization decays in the presence of two local motions. *Protein Science* **18**, 2003–2015 (2009).

[114] Ellis, J.P., Bakke, C.K., Kirchdoerfer, R.N., Jungbauer, L.M. & Cavagnero, S. Chain dynamics of nascent polypeptides emerging from the ribosome. *ACS Chemical Biology* **3**, 555–566 (2008).

[115] Kolb, V.A., Makeyev, E.V. & Spirin, A.S. Folding of firefly luciferase during translation in a cell-free system. *The EMBO journal* **13**, 3631–3637 (1994).

[116] Frydman, J., Nimmesgern, E., Ohtsuka, K. & Hartl, F.U. Folding of nascent polypeptide chains in a high molecular mass assembly with molecular chaperones. *Nature* **370**, 111–117 (1994).

[117] Svetlov, M.S. Effective cotranslational folding of firefly luciferase without chaperones of the Hsp70 family. *Protein Science* **15**, 242–247 (2006).

[118] Zhang, G. Personal Communication.

[119] Voss, N.R., Gerstein, M., Steitz, T.A. & Moore, P.B. The geometry of the ribosomal polypeptide exit tunnel. *J Mol Biol* **360**, 893–906 (2006).

[120] Wilson, D.N. & Beckmann, R. The ribosomal tunnel as a functional environment for nascent polypeptide folding and translational stalling. *Current opinion in chemical biology* **21**, 274–282 (2011).

[121] Clark, P.L. Protein folding in the cell: reshaping the folding funnel. *Trends in Biochemical Sciences* **29**, 527–534 (2004).

[122] Lavalette, D., Amand, B. & Pochon, F. Rotational relaxation of 70S ribosomes by a depolarization method using triplet probes. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 1407–1411 (1977).

[123] Schanda, P., Kupče, Ē. & Brutscher, B. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *Journal of Biomolecular NMR* **33**, 199–211 (2005).

[124] Cabrita, L.D., Hsu, S.T.D., Launay, H., Dobson, C.M. & Christodoulou, J. Probing ribosome-nascent chain complexes produced in vivo by NMR spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 22239–22244 (2009).

[125] Rutkowska, A., Beerbaum, M., Rajagopalan, N., Fiaux, J., Schmieder, P., Kramer, G., Oschkinat, H. & Bukau, B. Large-scale purification of ribosome-nascent chain complexes for biochemical and structural studies. *FEBS Letters* **583**, 2407–2413 (2009).

[126] Eichmann, C., Preissler, S., Riek, R. & Deuerling, E. Cotranslational structure acquisition of nascent polypeptides monitored by NMR spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 9111–9116 (2010).

[127] Wilkins, D.K., Grimshaw, S.B., Receveur, V., Dobson, C.M., Jones, J.A. & Smith, L.J. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* **38**, 16424–16431 (1999).

[128] Stejskal, E.O. & Tanner, J.E. Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient. *The Journal of Chemical Physics* **42**, 288–292 (1965).

[129] Tolman, J.R., Flanagan, J.M., Kennedy, M.A. & Prestegard, J.H. Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 9279–9283 (1995).

[130] Tjandra, N. & Bax, A. Direct Measurement of Distances and Angles in Biomolecules by NMR in a Dilute Liquid Crystalline Medium. *Science* **278**, 1111–1114 (1997).

[131] Sanders, C.R. & Landis, G.C. Facile Acquisition and Assignment of Oriented Sample NMR Spectra for Bilayer Surface-Associated Proteins. *Journal of the American Chemical Society* **116**, 6470–6471 (1994).

[132] Hansen, M.R., Mueller, L. & Pardi, A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature structural biology* **5**, 1065–1074 (1998).

[133] Clore, G.M., Starich, M.R. & Gronenborn, A.M. Measurement of Residual Dipolar Couplings of Macromolecules Aligned in the Nematic Phase of a Colloidal Suspension of Rod-Shaped Viruses. *Journal of the American Chemical Society* **120**, 10571–10572 (1998).

[134] Rückert, M. & Otting, G. Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments. *Journal of the American Chemical Society* **122**, 7793–7797 (2000).

[135] Sass, J., Cordier, F., Hoffmann, A., Rogowski, M., Cousin, A., Omichinski, J.G., Löwen, H. & Grzesiek, S. Purple Membrane Induced Alignment of Biological Macromolecules in the Magnetic Field. *Journal of the American Chemical Society* **121**, 2047–2055 (1999).

[136] Koenig, B.W., Hu, J.S., Ottiger, M., Bose, S., Hendler, R.W. & Bax, A. NMR Measurement of Dipolar Couplings in Proteins Aligned by Transient Binding to Purple Membrane Fragments. *Journal of the American Chemical Society* **121**, 1385–1386 (1999).

[137] Franciska, v.L. & Ruud, S. Media for orienting biological macromolecules. *Journal of Biomolecular NMR* pages 1–34 (2012).

[138] Bromley, K.M., Patil, A.J., Seddon, A.M., Booth, P. & Mann, S. Bio-Functional Mesolamellar Nanocomposites Based on Inorganic/Polymer Intercalation in Purple Membrane (Bacteriorhodopsin) Films. *Advanced Materials* **19**, 2433–2438 (2007).

[139] Nitz, M., Sherawat, M., Franz, K.J., Peisach, E., Allen, K.N. & Imperiali, B. Structural origin of the high affinity of a chemically evolved lanthanide-binding peptide. *Angewandte Chemie (International ed. in English)* **43**, 3682–3685 (2004).

[140] Banci, L., Bertini, I., Huber, J.G., Luchinat, C. & Rosato, A. Partial Orientation of Oxidized and Reduced Cytochrome b5at High Magnetic Fields: Magnetic Susceptibility Anisotropy Contributions and Consequences for Protein Solution Structure Determination. *Journal of the American Chemical Society* **120**, 12903–12909 (1998).

[141] Arnesano, F., Banci, L., Bertini, I., Van Der Wetering, K., Czisch, M. & Kaptein, R. The auto-orientation in high magnetic fields of oxidized cytochrome b562 as source of constraints for solution structure determination. *Journal of Biomolecular NMR* **17**, 295–304 (2000).

[142] Hus, J.C., Marion, D. & Blackledge, M. *De novo* determination of protein structure by NMR using orientational and long-range order restraints. *Current opinion in chemical biology* **298**, 927–936 (2000).

[143] Wöhnert, J., Franz, K.J., Nitz, M., Imperiali, B. & Schwalbe, H. Protein alignment by a coexpressed lanthanide-binding tag for the measurement of residual dipolar couplings. *Journal of the American Chemical Society* **125**, 13338–13339 (2003).

[144] Allen, J.E. & McLendon, G.L. Tryptophan and tyrosine to terbium fluorescence resonance energy transfer as a method to "map" aromatic residues and monitor docking. *Biochem Biophys Res Commun* **349**, 1264–1268 (2006).

[145] Su, X.C., Mcandrew, K., Huber, T. & Otting, G. Lanthanide-binding peptides for NMR measurements of residual dipolar couplings and paramagnetic effects from multiple angles. *Journal of the American Chemical Society* **130**, 1681–1687 (2008).

[146] Clore, G.M. & Iwahara, J. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chemical reviews* **109**, 4108–4139 (2009).

[147] Sculimbrene, B.R. & Imperiali, B. Lanthanide-binding tags as luminescent probes for studying protein interactions. *Journal of the American Chemical Society* **128**, 7346–7352 (2006).

[148] Thiriot, D.S. Structural basis of the temperature transition of Pf1 bacteriophage. *Protein Science* **14**, 1064–1070 (2005).

[149] Bothner-By, A.A. *Encyclopedia of Magnetic Resonance*. John Wiley & Sons, Ltd, Chichester, UK (2007).

[150] Zweckstetter, M. & Bax, A. Characterization of molecular alignment in aqueous suspensions of Pf1 bacteriophage. *Journal of Biomolecular NMR* **20**, 365–377 (2001).

[151] Hansen, M.R., Hanson, P. & Pardi, A. Filamentous bacteriophage for aligning RNA, DNA, and proteins for measurement of nuclear magnetic resonance dipolar coupling interactions. *Method Enzymol* **317**, 220–240 (2000).

[152] Ottiger, M. & Bax, A. Characterization of magnetically oriented phospholipid micelles for measurement of dipolar couplings in macromolecules. *Journal of Biomolecular NMR* **12**, 361–372 (1998).

[153] Cavagnero, S., Dyson, H.J. & Wright, P.E. Improved low pH bicelle system for orienting macromolecules over a wide temperature range. *Journal of Biomolecular NMR* **13**, 387–391 (1999).

[154] Prosser, R.S., Losonczi, J.A. & Shiyanovskaya, I.V. Use of a Novel Aqueous Liquid Crystalline Medium for High-Resolution NMR of Macromolecules in Solution. *Journal of the American Chemical Society* **120**, 11010–11011 (1998).

[155] Barrientos, L.G., Dolan, C. & Gronenborn, A.M. Characterization of surfactant liquid crystal phases suitable for molecular alignment and measurement of dipolar couplings. *Journal of Biomolecular NMR* **16**, 329–337 (2000).

[156] Tjandra, N. & Bax, A. Measurement of dipolar contributions to $^1J_{\mathrm{CH}}$ splittings from magnetic-field dependence of J modulation in two-dimensional NMR spectra. *Journal of magnetic resonance* **124**, 512–515 (1997).

[157] Ottiger, M., Delaglio, F. & Bax, A. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *Journal of magnetic resonance* **131**, 373–378 (1998).

[158] Hus, J.C., Marion, D. & Blackledge, M. Determination of Protein Backbone Structure Using Only Residual Dipolar Couplings. *Journal of the American Chemical Society* **123**, 1541–1542 (2001).

[159] Dyson, H.J. & Wright, P.E. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6**, 197–208 (2005).

[160] Uversky, V.N. Natively unfolded proteins: A point where biology waits for physics. *Protein Science* **11**, 739–756 (2002).

[161] Best, R.B. & Vendruscolo, M. Determination of Protein Structures Consistent with NMR Order Parameters. *Journal of the American Chemical Society* **126**, 8090–8091 (2004).

[162] Lindorff-Larsen, K., Best, R.B., DePristo, M.A., Dobson, C.M. & Vendruscolo, M. Simultaneous determination of protein structure and dynamics. *Nature* **433**, 128–132 (2005).

[163] Shortle, D. Persistence of Native-Like Topology in a Denatured Protein in 8 M Urea. *Science* **293**, 487–489 (2001).

[164] Louhivuori, M., Fredriksson, K., Pääkkönen, K., Permi, P. & Annila, A. Alignment of chain-like molecules. *Journal of Biomolecular NMR* **29**, 517–524 (2004).

[165] Louhivuori, M., Pääkkönen, K., Fredriksson, K., Permi, P., Lounila, J. & Annila, A. On the origin of residual dipolar couplings from denatured proteins. *Journal of the American Chemical Society* **125**, 15647–15650 (2003).

[166] Jha, A.K., Colubri, A., Freed, K.F. & Sosnick, T.R. Statistical coil model of the unfolded state: resolving the reconciliation problem. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 13099–13104 (2005).

[167] Bernado, P., Blanchard, L., Timmins, P., Marion, D., Ruigrok, R.W.H. & Blackledge, M. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 17002–17007 (2005).

[168] Bernado, P., Bertoncini, C.W., Griesinger, C., Zweckstetter, M. & Blackledge, M. Defining long-range order and local disorder in native α-synuclein using residual dipolar couplings. *Journal of the American Chemical Society* **127**, 17968–17969 (2005).

[169] Nodet, G., Salmon, L., Ozenne, V., Meier, S., Jensen, M.R. & Blackledge, M. Quantitative Description of Backbone Conformational Sampling of Unfolded Proteins at Amino Acid Resolution from NMR Residual Dipolar Couplings. *Journal of the American Chemical Society* **131**, 17908–17918 (2009).

[170] Huang, J.R. & Grzesiek, S. Ensemble calculations of unstructured proteins constrained by RDC and PRE data: a case study of urea-denatured ubiquitin. *Journal of the American Chemical Society* **132**, 694–705 (2010).

[171] Esteban-Martín, S., Fenwick, R.B. & Salvatella, X. Refinement of ensembles describing unstructured proteins using NMR residual dipolar couplings. *Journal of the American Chemical Society* **132**, 4626–4632 (2010).

[172] Bax, A. & Grishaev, A. Weak alignment NMR: a hawk-eyed view of biomolecular structure. *Current opinion in chemical biology* **15**, 563–570 (2005).

[173] Tugarinov, V. & Kay, L.E. Quantitative NMR Studies of High Molecular Weight Proteins: Application to Domain Orientation and Ligand Binding in the 723 Residue Enzyme Malate Synthase G. *Current opinion in chemical biology* **327**, 1121–1133 (2003).

[174] Velyvis, A., Yang, Y.R., Schachman, H.K. & Kay, L.E. A solution NMR study showing that active site ligands and nucleotides directly perturb the allosteric equilibrium in aspartate transcarbamoylase. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 8815–8820 (2007).

[175] Sprangers, R. & Kay, L.E. Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* **445**, 618–622 (2007).

[176] Sprangers, R. Quantitative NMR spectroscopy of supramolecular complexes: Dynamic side pores in ClpP are important for product release. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 16678–16683 (2005).

[177] Sprangers, R. & Kay, L.E. Probing supramolecular structure from measurement of methyl ^{1}H-^{13}C residual dipolar couplings. *Journal of the American Chemical Society* **129**, 12668–12669 (2007).

[178] Berk, V. & Cate, J.H.D. Insights into protein biosynthesis from structures of bacterial ribosomes. *Current opinion in chemical biology* **17**, 302–309 (2007).

[179] Mulder, F.A.A., Bouakaz, L., Lundell, A., Venkataramana, M., Liljas, A., Akke, M. & Sanyal, S. Conformation and dynamics of ribosomal stalk protein L12 in solution and on the ribosome. *Biochemistry* **43**, 5930–5936 (2004).

[180] Wahl, M.C., Bourenkov, G.P., Bartunik, H.D. & Huber, R. Flexibility, conformational diversity and two dimerization modes in complexes of ribosomal protein L12. *The EMBO journal* **19**, 174–186 (2000).

[181] Helgstrand, M., Mandava, C.S., Mulder, F.A.A., Liljas, A., Sanyal, S. & Akke, M. The ribosomal stalk binds to translation factors IF2, EF-Tu, EF-G and RF3 via a conserved region of the L12 C-terminal domain. *J Mol Biol* **365**, 468–479 (2007).

[182] Subramanian, A.R. Copies of proteins L7 and L12 and heterogeneity of the large subunit of *Escherichia coli* ribosome. *Current opinion in chemical biology* **95**, 1–8 (1975).

[183] Brimacombe, R., Stoffler, G. & Wittmann, H.G. Ribosome Structure. *Annu Rev Biochem* **47**, 217–249 (1978).

[184] Miyoshi, T., Nomura, T. & Uchiumi, T. Engineering and characterization of the ribosomal L10.L12 stalk complex. A structural element responsible for high turnover of the elongation factor G-dependent GTPase. *The Journal of Biological Chemistry* **284**, 85–92 (2009).

[185] Terhorst, C., Möller, W., Laursen, R. & Wittmann-Liebold, B. Amino acid sequence of a 50 S ribosomal protein involved in both EFG and EFT dependent GTP-hydrolysis. *FEBS Letters* **28**, 325–328 (1972).

[186] Möller, W.W., Groene, A.A., Terhorst, C.C. & Amons, R.R. 50-S ribosomal proteins. Purification and partial characterization of two acidic proteins, A1 and A2, isolated from 50-S ribosomes of *Escherichia coli*. *European journal of biochemistry / FEBS* **25**, 5–12 (1972).

[187] Kischa, K., Möller, W. & Stoffler, G. Reconstitution of a GTPase activity by a 50S ribosomal protein and *E. coli*. *Nature: New biology* **233**, 62–63 (1971).

[188] Gudkov, A., Tumanova, L., Gongadze, G. & Bushuev, V. Role of different regions of ribosomal proteins L7 and L10 in their complex formation and in the interaction with the ribosomal 50 S subunit. *FEBS Letters* **109**, 34 (1980).

[189] Korostelev, A., Trakhanov, S., Laurberg, M. & Noller, H.F. Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. *Cell* **126**, 1065–1077 (2006).

[190] Montesano-Roditis, L., Glitz, D.G., Traut, R.R. & Stewart, P.L. Cryo-electron microscopic localization of protein L7/L12 within the *Escherichia coli* 70S ribosome by difference mapping and Nanogold labeling. *The Journal of Biological Chemistry* **276**, 14117–14123 (2001).

[191] Datta, P.P., Sharma, M.R., Qi, L., Frank, J. & Agrawal, R.K. Interaction of the G' domain of elongation factor G and the C-terminal domain of ribosomal protein L7/L12 during translocation as revealed by cryo-EM. *Molecular Cell* **20**, 723–731 (2005).

[192] Traub, P. & NOMURA, M. Structure and functior of *Escherichia coli* ribosomes. *Current opinion in chemical biology* **34**, 575–593 (1968).

[193] Oleinikov, A.V., Jokhadze, G.G. & Traut, R.R. A single-headed dimer of *Escherichia coli* ribosomal protein L7/L12 supports protein synthesis. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 4215–4218 (1998).

[194] Mandava, C.S., Peisker, K., Ederth, J., Kumar, R., Ge, X., Szaflarski, W. & Sanyal, S. Bacterial ribosome requires multiple L12 dimers for efficient initiation and elongation of protein synthesis involving IF2 and EF-G. *Nucleic Acids Research* **40**, 2054–2064 (2012).

[195] Oleinikov, A.V., Perroud, B., Wang, B. & Traut, R.R. Structural and functional domains of *Escherichia coli* ribosomal protein L7/L12. The hinge region is required for activity. *The Journal of Biological Chemistry* **268**, 917–922 (1993).

[196] Gudkov, A.T., Budovskaya, E.V. & Sherstobaeva, N.M. The first 37 residues are sufficient for dimerization of ribosomal L7/L12 protein. *FEBS Letters* **367**, 280–282 (1995).

[197] Bocharov, E., Gudkov, A., Budovskaya, E. & Arseniev, A. Conformational independence of N-and C-domains in ribosomal protein L7/L12 and in the complex with protein L10. *FEBS Letters* **423**, 347–350 (1998).

[198] Bernado, P., Modig, K., Grela, P., Svergun, D.I., Tchorzewski, M., Pons, M. & Akke, M. Structure and Dynamics of Ribosomal Protein L12: An Ensemble Model Based on SAXS and NMR Relaxation. *Biophysical Journal* **98**, 2374–2382 (2010).

[199] Luer, C.A. & Wong, K.P. Conformation of *Escherichia coli* ribosomal protein L7/L12 in solution: hydrodynamic, spectroscopic, and conformation prediction studies. *Biochemistry* **18**, 2019–2027 (1979).

[200] Österberg, R., Sjöberg, B., Liljas, A. & Pettersson, I. Small-angle X-ray scattering and crosslinking study of the proteins L7/L12 from *Escherichia coli* ribosomes. *FEBS Letters* **66**, 48–51 (1976).

[201] Traut, R.R., Oleinikov, A.V., Makarov, E. & Jokhadze, G. Structure and function of *Escherichia coli* ribosomal protein L7/L12: effect of cross-links and deletions. *The translational apparatus: ...* (1993).

[202] Möller, W., Castleman, H. & Terhorst, C.P. Characterization of an acidic protein in 50S ribosomes of *E. coli*. *FEBS Letters* **8**, 192–196 (1970).

[203] Georgalis, Y., Dijk, J., Labischinski, H. & Wills, P.R. The tetrameric form of ribosomal protein L7/L12 from *Escherichia coli*. *The Journal of Biological Chemistry* **264**, 9210–9214 (1989).

[204] Kar, E.G. & Aune, K.C. Solution behavior of proteins L7/L12 from the 50S ribosomal subunit of *Escherichia coli*. *Biochemistry* **20**, 4638–4646 (1981).

[205] Hamman, B.D., Oleinikov, A.V., Jokhadze, G.G., Traut, R.R. & Jameson, D.M. Dimer/monomer equilibrium and domain separations of *Escherichia coli* ribosomal protein L7/L12. *Biochemistry* **35**, 16680–16686 (1996).

[206] Launay, H.M.M. Personal Communication.

[207] Launay, H.M.M. NMR investigations of co-translational protein folding on ribosomal particles. *PhD Thesis* pages 1–244 (2012).

[208] Ortega, A., Amorós, D. & García de la Torre, J. Prediction of Hydrodynamic and Other Solution Properties of Rigid Proteins from Atomic- and Residue-Level Models. *Biophysical Journal* **101**, 892–898 (2011).

[209] Kontaxis, G., Clore, G. & Bax, A. Evaluation of cross-correlation effects and measurement of one-bond couplings in proteins with short transverse relaxation times. *Journal of magnetic resonance* **143**, 184–196 (2000).

[210] Lipari, G. & Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *Journal of the American Chemical Society* **104**, 4546–4559 (1982).

[211] Bocharov, E.V., Gudkov, A.T. & Arseniev, A.S. Topology of the secondary structure elements of ribosomal protein L7/L12 from *E. coli* in solution. *FEBS Letters* **379**, 291–294 (1996).

[212] Vallurupalli, P., Hansen, D.F., Stollar, E., Meirovitch, E. & Kay, L.E. Measurement of bond vector orientations in invisible excited states of proteins. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 18473–18477 (2007).

[213] Zweckstetter, M. NMR: prediction of molecular alignment from structure using the PALES software. *Nat Protoc* **3**, 679–690 (2008).

[214] Kirkpatrick, J.P., Li, P. & Carlomagno, T. Probing Mutation-Induced Structural Perturbations by Refinement Against Residual Dipolar Couplings: Application to the U4 Spliceosomal RNP Complex. *ChemBioChem* **10**, 1007–1014 (2009).

[215] Schwieters, C.D., Kuszewski, J.J., Tjandra, N. & Clore, G.M. The Xplor-NIH NMR molecular structure determination package. *Journal of magnetic resonance* **160**, 65–73 (2003).

[216] Ottiger, M., Delaglio, F., Marquardt, J.L., Tjandra, N. & Bax, A. Measurement of dipolar couplings for methylene and methyl sites in weakly oriented macromolecules and their use in structure determination. *Journal of magnetic resonance* **134**, 365–369 (1998).

[217] Giesen, A.W., Bae, L.C., Barrett, C.L., Chyba, J.A., Chaykovsky, M.M., Cheng, M.C., Murray, J.H., Oliver, E.J., Sullivan, S.M., Brown, J.M., Dahlquist, F.W. & Homans, S.W. Measurement of one-bond ^{1}H-^{13}C, couplings in backbone-labelled proteins. *Journal of Biomolecular NMR* **19**, 255–260 (2001).

[218] Permi, P. Measurement of residual dipolar couplings from $^{1}H^{\alpha}$ to $^{13}C^{\alpha}$ and ^{15}N using a simple HNCA-based experiment. *Journal of Biomolecular NMR* **27**, 341–349 (2003).

[219] Yang, D., Tolman, J.R., Goto, N.K. & Kay, L. An HNCO-based Pulse Scheme for the Measurement of $^{13}C^{\alpha}$-$^{1}H^{\alpha}$ One-bond Dipolar couplings in ^{15}N, ^{13}C Labeled Proteins. *Journal of Biomolecular NMR* **12**, 325–332 (1998).

[220] Hitchens, T.K., McCallum, S.A. & Rule, G.S. A J^{CH}-modulated 2D (HACACO)NH pulse scheme for quantitative measurement of $^{13}C^{\alpha}$-$^{1}H^{\alpha}$ couplings in ^{15}N, ^{13}C-labeled proteins. *Journal of magnetic resonance* **140**, 281–284 (1999).

[221] Chou, J.J. & Bax, A. Protein Side-Chain Rotamers from Dipolar Couplings in a Liquid Crystalline Phase. *Journal of the American Chemical Society* **123**, 3844–3845 (2001).

[222] de Alba, E., Suzuki, M. & Tjandra, N. Simple multidimensional NMR experiments to obtain different types of one-bond dipolar couplings simultaneously. *Journal of Biomolecular NMR* **19**, 63–67 (2001).

[223] Tjandra, N. & Bax, A. Large Variations in $^{13}C^{\alpha}$ Chemical Shift Anisotropy in Proteins Correlate with Secondary Structure. *Journal of the American Chemical Society* **119**, 9576–9577 (1997).

[224] Lescop, E., Schanda, P. & Brutscher, B. A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *Journal of magnetic resonance* **187**, 163–169 (2007).

[225] Hsu, S.T.D., Fucini, P., Cabrita, L.D., Launay, H., Dobson, C.M. & Christodoulou, J. Structure and dynamics of a ribosome-bound nascent chain by NMR spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 16516–16521 (2007).

[226] Tugarinov, V., Hwang, P.M., Ollerenshaw, J.E. & Kay, L.E. Cross-correlated relaxation enhanced ^{1}H-^{13}C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *Journal of the American Chemical Society* **125**, 10420–10428 (2003).

[227] Williamson, M.P., Havel, T.F. & Wüthrich, K. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by ^{1}H nuclear magnetic resonance and distance geometry. *J Mol Biol* **182**, 295–315 (1985).

[228] Gilbert, R.J.C., Fucini, P., Connell, S., Fuller, S.D., Nierhaus, K.H., Robinson, C.V., Dobson, C.M. & Stuart, D.I. Three-dimensional structures of translating ribosomes by Cryo-EM. *Mol Cell* **14**, 57–66 (2004).

[229] Schaffitzel, C. & Ban, N. Reprint of "Generation of ribosome nascent chain complexes for structural and functional studies" [J. Struct. Biol. 158 (2007) 463-471]. *J Struct Biol* **159**, 302–310 (2007).

[230] Evans, M.S., Ugrinov, K.G., Frese, M.A. & Clark, P.L. Homogeneous stalled ribosome nascent chain complexes produced in vivo or in vitro. *Nature methods* **2**, 757–762 (2005).

[231] Nakatogawa, H. & Ito, K. The ribosomal exit tunnel functions as a discriminating gate. *Cell* **108**, 629–636 (2002).

[232] Ito, K., Chiba, S. & Pogliano, K. Divergent stalling sequences sense and control cellular physiology. *Biochem Biophys Res Commun* **393**, 1–5 (2010).

[233] Nakatogawa, H., Murakami, A. & Ito, K. Control of SecA and SecM translation by protein secretion. *Curr Opin Microbiol* **7**, 145–150 (2004).

[234] Vázquez-Laslop, N., Ramu, H., Klepacki, D., Kannan, K. & Mankin, A.S. The key function of a conserved and modified rRNA residue in the ribosomal response to the nascent peptide. *The EMBO journal* **29**, 3108–3117 (2010).

[235] McCoy, A.J., Fucini, P., Noegel, A.A. & Stewart, M. Structural basis for dimerization of the Dictyostelium gelation factor (ABP120) rod. *Nature structural biology* **6**, 836–841 (1999).

[236] Popowicz, G.M., Müller, R., Noegel, A.A., Schleicher, M., Huber, R. & Holak, T.A. Molecular structure of the rod domain of dictyostelium filamin. *J Mol Biol* **342**, 1637–1646 (2004).

[237] Hsu, S.T.D., Cabrita, L.D., Christodoulou, J. & Dobson, C.M. ^{1}H, ^{15}N and ^{13}C assignments of domain 5 of Dictyostelium discoideum gelation factor (ABP-120) in its native and 8M urea-denatured states. *Biomolecular NMR Assignments* **3**, 29–31 (2009).

[238] Studier, F.W. Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification* **41**, 207–234 (2005).

[239] Calvori, C., Frontali, L., Leoni, L. & Tecce, G. Effect of rifamycin on protein synthesis. *Nature* **207**, 417–418 (1965).

[240] Campbell, E.A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A. & Darst, S.A. Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase. *Current opinion in chemical biology* **104**, 901–912 (2001).

[241] Feklistov, A., Mekler, V., Jiang, Q., Westblade, L.F., Irschik, H., Jansen, R., Mustaev, A., Darst, S.A. & Ebright, R.H. Rifamycins do not function by allosteric modulation of binding of Mg2+ to the RNA polymerase active center. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 14820–14825 (2008).

[242] Cabrita, L.D. Personal communication.

[243] Maroteaux, L., Campanelli, J.T. & Scheller, R.H. Synuclein: a neuron-specific protein localized to the nucleus and presynaptic nerve terminal. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **8**, 2804–2815 (1988).

[244] Jakes, R., Spillantini, M.G. & Goedert, M. Identification of two distinct synucleins from human brain. *FEBS Letters* **345**, 27–32 (1994).

[245] Goedert, M. Alpha-synuclein and neurodegenerative diseases. *Nature reviews. Neuroscience* **2**, 492–501 (2001).

[246] Ulmer, T.S. Structure and Dynamics of Micelle-bound Human -Synuclein. *The Journal of Biological Chemistry* **280**, 9595–9603 (2004).

[247] Giasson, B.I., Murray, I.V., Trojanowski, J.Q. & Lee, V.M. A hydrophobic stretch of 12 amino acid residues in the middle of α-synuclein is essential for filament assembly. *The Journal of Biological Chemistry* **276**, 2380–2386 (2001).

[248] Dedmon, M.M., Lindorff-Larsen, K., Christodoulou, J., Vendruscolo, M. & Dobson, C.M. Mapping Long-Range Interactions in -Synuclein using Spin-Label NMR and Ensemble Molecular Dynamics Simulations. *Journal of the American Chemical Society* **127**, 476–477 (2005).

[249] Croke, R.L., Sallum, C.O., Watson, E., Watt, E.D. & Alexandrescu, A.T. Hydrogen exchange of monomeric α-synuclein shows unfolded structure persists at physiological temperature and is independent of molecular crowding in *Escherichia coli*. *Protein Science* **17**, 1434–1445 (2008).

[250] Ottiger, M. & Bax, A. Determination of Relative N-HN, N-C', Ca -C', and Ca -Ha Effective Bond Lengths in a Protein by NMR in a Dilute Liquid Crystalline Phase. *Journal of the American Chemical Society* **120**, 12334–12341 (1998).

[251] Maltsev, A.S., Ying, J. & Bax, A. Impact of N-terminal acetylation of α-synuclein on its random coil and lipid binding properties. *Biochemistry* **51**, 5004–5013 (2012).

[252] Hautbergue, G.M. & Golovanov, A.P. Increasing the sensitivity of cryoprobe protein NMR experiments by using the sole low-conductivity arginine glutamate salt. *Journal of magnetic resonance* **191**, 335–339 (2008).

[253] Hsu, S.T.D., Cabrita, L.D., Fucini, P., Christodoulou, J. & Dobson, C.M. Probing side-chain dynamics of a ribosome-bound nascent chain using methyl NMR spectroscopy. *Journal of the American Chemical Society* **131**, 8366–8367 (2009).

[254] Farès, C., Lakomek, N.A., Walter, K.F.A., Frank, B.T.C., Meiler, J., Becker, S. & Griesinger, C. Accessing ns-micros side chain dynamics in ubiquitin with methyl RDCs. *Journal of Biomolecular NMR* **45**, 23–44 (2009).

[255] Liao, X., Godoy-Ruiz, R., Guo, C. & Tugarinov, V. Simultaneous measurement of ^{1}H-^{15}N and methyl $^{1}H_{m}$-$^{13}C_{m}$ residual dipolar couplings in large proteins. *Journal of Biomolecular NMR* **51**, 191–198 (2011).

[256] Sivashanmugam, A., Murray, V., Cui, C., Zhang, Y., Wang, J. & Li, Q. Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Science* **18**, 936–948 (2009).

[257] Hill, W.E., Rossetti, G.P. & Van Holde, K.E. Physical studies of ribosomes from *Escherichia coli*. *J Mol Biol* **44**, 263–277 (1969).

[258] Rasband, W. Image J .

[259] Hwang, T.L. & Shaka, A.J. Water suppression that works. Excitation sculpting using arbitrary wave-forms and pulsed-field gradients. *Journal of magnetic resonance* **112**, 275–279 (1995).

[260] Schanda, P. & Brutscher, B. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *Journal of the American Chemical Society* **127**, 8014–8015 (2005).

[261] Shaka, A.J., Barker, P.B. & Freeman, R. Computer-optimized decoupling scheme for wideband applications and low-level operation. *Journal of magnetic resonance* **64**, 547–552 (1985).

[262] Kupce, E. & Freeman, R. Polychromatic selective pulses. *Journal of magnetic resonance* **102**, 122–126 (1993).

[263] Geen, H. & Freeman, R. Band-selective radiofrequency pulses. *Journal of magnetic resonance* **93**, 93–141 (1991).

[264] Fu, R. & Bodenhausen, G. Broadband decoupling in NMR with frequency-modulated 'chirp'pulses. *Chemical Physics Letters* **245**, 415–420 (1995).

[265] Tanner, J.E. Use of the stimulated echo in NMR diffusion studies. *The Journal of Chemical Physics* **52**, 2523 (1970).

[266] Augustyniak, R., Ferrage, F., Paquin, R., Lequin, O. & Bodenhausen, G. Methods to determine slow diffusion coefficients of biomolecules. Applications to Engrailed 2, a partially disordered protein. *Journal of Biomolecular NMR* **50**, 209–218 (2011).

[267] Bodenhausen, G. & Ruben, D. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters* **69**, 185–189 (1980).

[268] Piotto, M., Saudek, V. & Sklenar, V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *Journal of Biomolecular NMR* **2**, 661–665 (1992).

[269] Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 12366–12371 (1997).

[270] Sklenar, V., Piotto, M., Leppik, R. & Saudek, V. Gradient-tailored water suppression for ^{1}H-^{15}N HSQC experiments optimized to retain full sensitivity. *Journal of magnetic resonance* **102**, 241–245 (1993).

[271] Emsley, L. & Bodenhausen, G. Optimization of shaped selective pulses for NMR using a quaternion description of their overall propagators. *Journal of magnetic resonance* **97**, 135–148 (1992).

[272] Smith, M.A., Hu, H. & Shaka, A.J. Improved Broadband Inversion Performance for NMR in Liquids. *Journal of magnetic resonance* **151**, 269–283 (2001).

[273] Grzesiek, S. & Bax, A. Amino acid type determination in the sequential assignment procedure of uniformly ^{13}C/^{15}N-enriched proteins. *Journal of Biomolecular NMR* **3**, 185–204 (1993).

[274] Santoro, J. & King, G.C. A constant-time 2D overbodenhausen experiment for inverse correlation of isotopically enriched species. *Journal of magnetic resonance* **97**, 202–207 (1992).

[275] Kay, L., Torchia, D.A. & Bax, A. Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28**, 8972–8979 (1989).

[276] Dayie, K.T. & Wagner, G. Relaxation-rate measurements for ^{15}N-^{1}H groups with pulsed-field gradients and preservation of coherence pathways. *Journal of magnetic resonance* **111**, 121–126 (1994).

[277] Kay, L., Nicholson, L.K., Delaglio, F., Bax, A. & Torchia, D.A. Pulse sequences for removal of the effects of cross correlation between dipolar and chemical-shift anisotropy relaxation mechanisms on the measurement of heteronuclear T1 and T2 values in proteins. *Journal of magnetic resonance* **97**, 359–375 (1992).

[278] Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D. & Kay, L. Backbone dynamics of a free and a phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry* **33**, 5984–6003 (1994).

[279] Korzhnev, D.M., Skrynnikov, N.R., Millet, O., Torchia, D.A. & Kay, L.E. An NMR Experiment for the Accurate Measurement of Heteronuclear Spin-Lock Relaxation Rates. *Journal of the American Chemical Society* **124**, 10743–10753 (2002).

[280] Hansen, D.F. & Kay, L.E. Improved magnetization alignment schemes for spin-lock relaxation experiments. *Journal of Biomolecular NMR* **37**, 245–255 (2007).

[281] Li, Y.C. & Montelione, G.T. Overcoming solvent saturation-transfer artifacts in protein NMR at neutral pH. Application of pulsed field gradients in measurements of ^{1}H-^{15}N Overhauser effects. *Journal of magnetic resonance* **105**, 45–51 (1994).

[282] Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. & Bax, A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **6**, 277–293 (1995).

[283] Boucher, W. Azara .

[284] Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J. & Laue, E.D. The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins* **59**, 687–696 (2005).

[285] Dosset, P., Hus, J.C., Blackledge, M. & Marion, D. Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *Journal of Biomolecular NMR* **16**, 23–28 (2000).

[286] Losonczi, J.A., Andrec, M., Fischer, M.W. & Prestegard, J.H. Order matrix analysis of residual dipolar couplings using singular value decomposition. *Journal of magnetic resonance* **138**, 334–342 (1999).

[287] Zweckstetter, M. & Bax, A. Evaluation of uncertainty in alignment tensors obtained from dipolar couplings. *Journal of Biomolecular NMR* **23**, 127–137 (2002).

[288] Delano, W.L. The PyMOL Molecular Graphics System .